Cross-species Microarray Analysis of Limb Development in the Bat, *Miniopterus natalensis*

Mandy K. Mason

Department of Zoology

University of Cape Town

Rondebosch

7701

Supervisors:

Prof. David Jacobs

Prof. Nicola Illing

February 2009

This dissertation was conducted in fulfilment of the requirements of a Masters of Science Degree in Zoology, MSc (Zoo), at the University of Cape Town.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Declaration

I, Mandy Mason, know the meaning of plagiarism and declare that all of the work in the document, save for that which is properly acknowledged, is my own.

Signed: Signed by candidate Date: 29th May 2009

Acknowledgements

I would like to thank my supervisors, Prof. David Jacobs and Assoc. Prof. Nicola Illing, who played a pivotal role in the concept, financing, support and completion of this study. Funding awards, the KW Johnstone Research Scholarship and the Harry Crossley Foundation Postgraduate Bursary, were provided by the UCT Postgraduate Funding Office and a Prestigious Award by the National Research Foundation. Thanks also go to my colleague in bat research, Dorit Hockman, and my partner, Robyn Verrinder who played integral roles in this project, both in the field and the lab. I must also thank Prof. Richard Behringer and Dr. Chris Cretekos for their collaboration and candid support of our fledgling bat developmental research team. Thanks go to Prof. Brigid Hogan for being a willing and capable participant in the field, for her generous gift of dissection equipment and for teaching us how to use them. My gratitude goes to Johan van Heerden, Dr. Sally-Ann Walford and Arthur Shen for giving me advice on all aspects of microarray experimentation and to Lab 425 for their support. Much needed technical support was also given by Faezah Davids, Petra Muller, Jaap Visser, and Hiram Arendse. Thanks go to Cape Nature and, more specifically, to the De Hoop Nature Reserve Manager, Peter Chadwick and the reserve staff for allowing us access, accommodation and logistical support when out in the field. I must also thank Stacey Jordan and the third year project teams for their dedicated work on projects associated with this study. Finally, thanks must go to the Bat Lab members Dr. Samantha Stoffberg, Lizelle Odendaal and Elizabeth Kelly for all the advice, support and encouragement. And thanks also to my family, for putting up with me during this time.

Abstract

This study reports the first characterisation of the embryonic bat limb transcriptome, allowing the identification of novel candidate genes that were differentially expressed between the bat hand and foot plate. These genes may have played important roles in the evolution of the bat wing and hindlimb. The reproduction and development of an African bat species, *Miniopterus natalensis*, was characterised and three maternal features (mass, belly size and plasma progesterone levels) examined as potential predictors of embryonic stages of development. Belly palpitation was found to be a useful field method to distinguish between non-pregnancy, early development or late development in female bats. A microarray analysis between the hand and foot plates of CS16 and CS17 bat embryos, and the hand plates of E13.5 mouse embryos, revealed high correlation between the transcriptomes of the bat autopods and the mouse hand plate ($r > 0.88$) and among all the bat autopods ($r > 0.98$). However, ten genes were found to be differentially expressed in both the CS16 and the CS17 bat hand plate as compared to the mouse hand plate while only three genes were identified as being significantly differentially expressed between bat foot plates and mouse hand plates. A comparison between the bat hand and foot plates identified fifteen genes that were differentially expressed at the stage CS17 stage and six at the stage CS16. Closer examination of gene families involved in limb development revealed novel expression of genes in the retinoic acid (RA) pathway, and the *Hoxd* family. This included the apparent co-regulation of the 5′ *Hoxd* genes (*Hoxd10*, *11*, *12* and *13*). Of the genes characterised in bat limb development (*Hoxd13*, *Bmp2*, *Fgf8* and *Prrx1*), higher mRNA transcript levels in the CS17 bat hand plate relative to the mouse hand plate was found for *Hoxd13* (FC = 2.6) and *Prrx1* (FC = 1.8). These differences were also found in a comparison to the CS17 bat foot plate (*Hoxd13*: FC = 1.4; *Prrx1*: FC = 1.4). A potentially novel transcript of *Meis2*, a gene important in specifying the proximal-distal (P-D) axis of the limb, was noted for its high fold changes in the bat hand plate as compared the foot plate (CS17: FC = 7.0; CS16: FC = 2.2) and the mouse hand plate (FC = 13.1).

Table of Contents

List of Illustrations

FIGURES:

TABLES:

Glossary and Abbreviations

A values	Average intensity data
aaUTP	P3-(4-azidoanilido)uridine 5′-triphosphate
AER	Apical ectodermal ridge
ALDH	Aldehyde dehydrogenases
A-P	anterior-posterior
aRNA	Amplified RNA
Aspect ratio	The ratio of wing length to wing area used to describe the shape of the bat wing
Bmp2	*Bone morphogenetic protein 2*
Bmp4	*Bone morphogenetic protein 4*
Bmp7	*Bone morphogenetic protein 7*
Bmp10	*Bone morphogenetic protein 10*
Bmpr1b	*Bone morphogenetic protein receptor, type 1B*
BSA	Bovine serum albumin
cDNA	Complementary DNA
Chiropatagium	The membrane structures of the bat wing that joins the digits together
COMBAT	Combating batch effects
CPGR	Centre for Proteomic and Genomic Research
CR	Count ratio
CRL	Crown Rump Length
CRABP	Cellular retinoic acid binding proteins
Crabp1	*Cellular retinoic acid binding protein 1*
CS16	*Carollia* stage 16
CS17	*Carollia* stage 17
CSH	Cross-species hybridisation
CTS	Cross-strain hybridisation
Cy3	Cyanine 3
Cy5	Cyanine 5
Cyp26a1	*Cytochrome P450, family 26, subfamily a, polypeptide 1*
DEDS	Differential expression via distance synthesis
DEPC	Diethylene pyrocarbonate
DNA	Deoxyribonucleic acid
dpc	Days post coitus
dsDNA	Double stranded DNA
D-V	Dorso-ventrally
E11.5	Embryonic day 11.5
E12.5	Embryonic day 12.5

E13.5	Embryonic day 13.5
Epha3	Ephrin receptor A3
Evo-devo	Evolutionary Development
FC	Fold change
FDR	False discovery rate
Fgf8	*Fibroblast growth factor 8*
Fgf9	*Fibroblast growth factor 9*
Fgf18	*Fibroblast growth factor 18*
Fgfr2	*Fibroblast growth factor receptor 2*
Foxk2	*Forkhead box K2*
FP	Foot plate
Fzdb	*Frizzled related protein-b1*
GCR	Global control region
Gdf10	*Growth differentiation factor 10*
Gfra2	*Glial cell line derived neurotrophic factor family receptor alpha 2*
GO	Gene ontology
Hapln1	*Hyaluronan and proteoglycan link protein 1*
Heterochrony	Developmental change in event timing
Hibernacula	Cool, humid roosts that are used by colonies of bats as shelter during periods of winter hibernations
Hip	Hedgehog-interacting protein
Homeodomain	A domain in a protein that is encoded for by a homeobox, that consists of about 60 amino acid residues which are usually similar from one such domain to another, and that recognizes and binds to specific DNA sequences in genes regulated by the homeotic gene
Hoxa9	*Homeobox A9*
Hoxa10	*Homeobox A10*
Hoxa11	*Homeobox A11*
Hoxa13	*Homeobox A13*
Hoxc6	*Homeobox C6*
Hoxc10	*Homeobox C10*
Hoxd9	*Homeobox D9*
Hoxd10	*Homeobox D10*
Hoxd11	*Homeobox D11*
Hoxd12	*Homeobox D12*
Hoxd13	*Homeobox D13*
HP	Hand plate
Igf-1a	Insulin-like growth factor IA precursor
Igf-1b	Insulin-like growth factor IB precursor
IVT	*In vitro* transcription
KNN	k- nearest neighbour

LPM	Lateral plate mesoderm
Ltbp4	*Latent transforming growth factor beta binding protein 4*
M values	Log ratio data
Maternity roosts	Warm, humid roosts that are used by colonies of bats as shelter during their reproductive periods
Meis1	*Meis homeobox 1*
Meis2	*Meis homeobox 2*
Meis3	*Meis homeobox 3*
µg	Micrograms
mm	Millimetres
mod*t*	Moderated *t*
mRNA	Messenger ribonucleic acid
Msx1	*Homeobox, msh-like 1*
Msx2	*Homeobox, msh-like 2*
NaOH	Sodium Hydroxide
NCBI	National Centre for Biotechnology Information
Nfib	*Nuclear factor I/B*
nm	Nanometres
NP	Not pregnant
NRMS	Normalised root mean square
Odz4	*Odd Oz/ten-m homolog 4 (Drosophila)*
oligo (dT) primer	A single strand of poly T bases that contain a T7 promoter
oligonucleotides	Short sections of DNA with a sequence that corresponds to that of a specific gene
Pbx1	*Pre B-cell leukaemia transcription factor 1*
Pbx2	*Pre B-cell leukaemia transcription factor 2*
Pbx3	*Pre B-cell leukaemia transcription factor 3*
PCA	Principal components analysis
P-D	Proximal-distal
P-D	proximal-distal
PFA	Paraformaldehyde
Phocomelia	A phenotype of shortened long bones and limbs
Pitx1	*Paired-like homeodomain transcription factor 1*
Plagiopatagium	The largest membrane structure of the bat wing. It joins the hind and the forelimb and is extended along the outer edge of the fifth digit
pmol/µl	Picomols per microlitre
PMT	Photomultiplier tube
polyA	Polyadenosine
Primordium	An organ or tissue in its earliest recognizable stage of development
Protopatagium	The small membrane structure of the bat wing that extends from the shoulder of the bat to its wrist

Prrx1	*Paired related homeobox 1*
Prrx2	*Paired related homeobox 2*
Ptk2	*Protein tyrosine kinase 2*
θ	Normalised root mean square error
RA	Retinoic acid
Rarβ	*Retinoic acid receptor, beta*
Rarres2	*Retinoic acid receptor responder (tazarotene induced) 2*
RBP	Retinol binding proteins
RDH	Retinol dehydrogenase
Rdh10	*Retinol dehydrogenase 10*
Rdh11	*Retinol dehydrogenase 11*
Rdh14	*Retinol dehydrogenase 14*
Rdh9	*Retinol dehydrogenase 9*
RIN	RNA integrity number
RNA	Ribonucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal RNA
RT-PCR	Real time polymerase chain reaction
SAM	Significance analysis of microarrays
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SE	Standard error
Sgk	*Serum/glucocorticoid regulated kinase 1*
SSC	Sodium chloride sodium citrate
SSH	Species-specific hybridisations
Tbx4	*TATA-box 4*
Tbx5	*TATA-box 4*
Tbx18	*TATA-box18*
TGF-betas	Transforming growth factor beta proteins
Tnc	*Tenascin C*
UTR	Untranslated region
Uropatagium	The membrane structure of the bat that stretches between the tail and the hindlimb skeletal elements
V	Volt
WT	Wild-type

Chapter 1

Evolution of the bat wing

1.1 THE EVOLUTIONARY ENIGMA OF BAT FLIGHT

The bat wing is a unique evolutionary structure. Its distal portion is homologous to the human hand (hence the name Chiroptera, or 'hand-wing' for the order), however it has distinctive morphological features, such as highly elongated digits and extensive interdigital membranes. These wings give bats the ability of highly manoeuvrable, powered flight, a feature that has allowed this mammal to radiate into the diverse order that it is today (Hill and Smith, 1984). There has been a long-standing interest in how such a morphologically distinct structure could have evolved. Darwin (1890) himself found it difficult to imagine how natural selection could have led to the adaptation of flight, as it is thought that any intermediate, flightless form would have been a disadvantageous one. The absence of these theoretical intermediate forms in the fossil record has led some to suggest that, on an evolutionary scale, the development of the bat wing may have been a relatively rapid event that arose only once in this order (Sears et al., 2006; Simmons and Geisler, 1998). This feature makes the bat wing a useful model for examining evolutionary mechanisms on a molecular scale (Sears, 2008; Sears et al., 2006)

1.2 THE BAT WING AS A UNIQUE STRUCTURE

The order Chiroptera is the second largest of all the mammals, it is composed of over a thousand species, most of which have adapted to distinct environmental conditions (Simmons, 2005). The bat wing is a thin, membranous structure that is supported by the forelimb, hindlimb and tail skeletal elements (Fig. 1.1A). The forearm and digital skeletal elements are disproportionately elongated, which allow the extended membranous wing to have a very large surface area. In an analogy given by Neuweiler (2000), the wing structure of the bat can be compared to that of an umbrella, where the firm skeletal 'spokes' support and hold the membrane rigid and taught or allow it to collapse and fold away neatly against the body. This wing structure, composed of flexible bones, membrane and independently controllable joints gives the bat unique flight capabilities (Swartz et al., 2006).

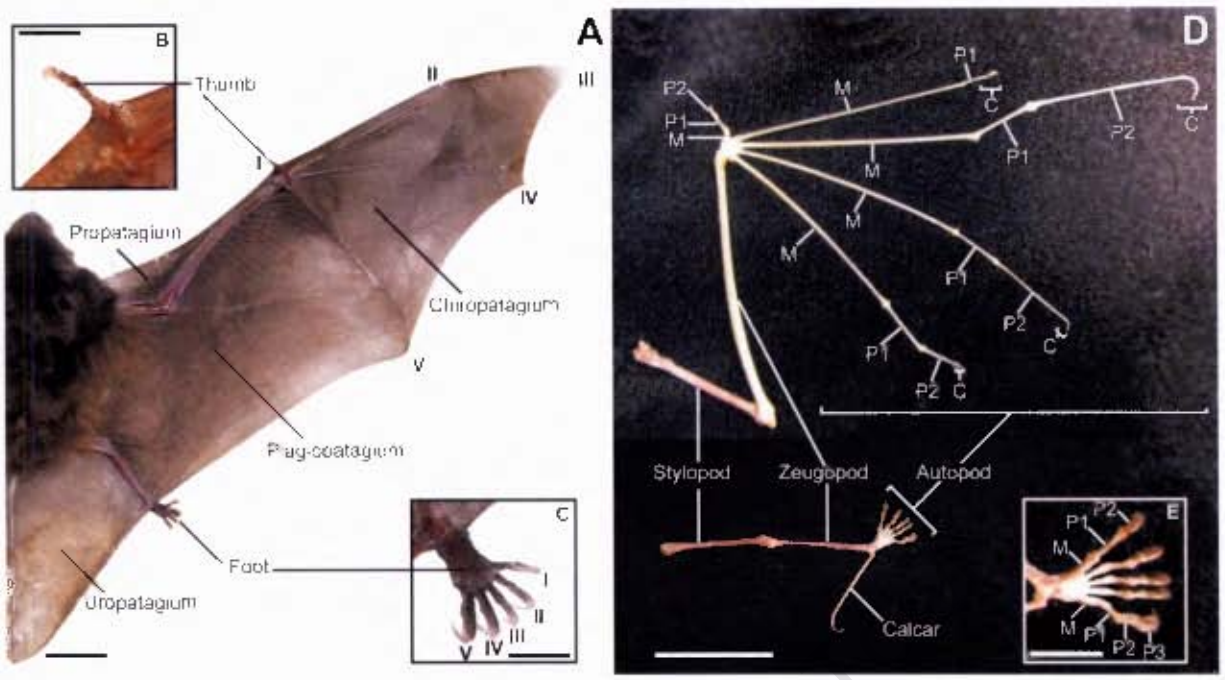


Fig. 1.1: Ventral view of the left side of an adult *Miniopterus natalensis* with wing membranes indicated (A). The thumb (B) and the digits of the foot (C) are similar in morphology and size. A dorsal view of the right forelimb and hindlimb skeletal elements (D) shows that the stylopod, the zeugopod and the metacarpals (M) and phalanges (P) of digits 2 to 5 of the forelimb are elongated in comparison to those of the foot (E). The distal tips of these elements retain cartilage (C). Scale bars A & D – 2 cm; B, C & E – 0.5 cm.

The hand-wing of the bat is composed of six carpals (wrist bones), which are in turn, joined to five metacarpals (hand bones) and five sets of phalanges (finger bones). The second to the fifth metacarpal and phalange are greatly elongated while; those of the first (the thumb) are not (Fig. 1.1D). The thumb functions as a freely moving grasping tool and to this end it has retained a claw while in most of the other digits this feature has been lost (Fig. 1.1B). The bat forelimb skeletal elements are significantly longer than those found in other (non-flying) mammals (Swartz, 1999). The second to the fifth digits have been integrated into the membranous wing. This is shown to have four different structural areas, collectively known as patagia. They are the pro-; chiro-; plagio-; and the uropatagium (Fig. 1.1A). Each of these membrane structures is derived from different embryonic primordia (Cretekos et al., 2005). The plagiopatagium is the largest wing membrane structure. It joins the hindlimb to the forelimb and is extended along the outer edge of the fifth digit. The protopatagium is a small area of membrane that extends from the shoulder of the bat to its wrist. The plagiopatagium and protopatagium support the weight of the bat's body during flight. The propatagium also acts as an adjustable leading edge for the wing to control lift (Norberg, 1972). The chiropatagium is the membrane that joins the digits together. It functions to propel the bat forward during flight. The uropatagium is the membrane that stretches between the tail and

the hindlimb skeletal elements. During flight this membrane can be extended to act as a breaking mechanism, some bats also use this structure to scoop up prey (Neuweiler, 2000).

The legs of the bat are also morphologically unique structures. They are small, thin and have little musculature. They have been found to be proportionally smaller than the hindlimb skeletal elements of other (non-flying) mammals (Swartz, 1999). The foot is uniform with little difference between digits, the exception being digit 1 (the big toe) which has only one joint as opposed to two in the other digits (Fig. 1.1E). There is no webbing between the digits of the foot (Fig. 1.1C). During flight the legs of the bat serve as a support structure for the uro- and plagiopatagium. The legs are adapted for grasping, they have retained their claws and allow bats to hang upside down when not in flight (Neuweiler, 2000)

There is a great diversity of bat species and a correspondingly large range of wing shapes and sizes among these species. Differences in wings among bat species give these species distinct aerodynamic abilities. In a very simplistic model, broad, short wings give the bats the ability of slow, manoeuvrable flight, while long, narrow wings allow for speed with less manoeuvrability (Norberg and Rayner, 1987). Bats that hunt above vegetation are more likely to be slim, long-winged, fast fliers; those that forage near or within vegetation are likely to be broad, short-winged, manoeuvrable fliers (Norberg and Rayner, 1987). This has allowed species to become well adapted to certain environmental niches.

1.3 THE EVOLUTION OF FLIGHT IN BATS

Bats appear fully developed in the fossil record of the early Eocene period (49-53 million years ago), without any record of an intermediate species (Simpson, 1953; Speakman, 2001). The majority of extinct bats show no significant differences in their forelimb digit lengths and widths to those of modern bats, indicating that these fossil bats were capable of powered flight (Sears et al., 2006). The exception is the earliest described fossil bat, *Onychonycteris finneyi*, which, had limb proportions intermediate between those found in bats and those found in forelimb dominated mammals (Simmons et al., 2008). These bats were still capable of powered flight as evidenced by the presence of other distinct flight associated skeletal features. However, due to the relatively small size of their cochlea, it is thought that they were unable to echolocate. This finding supports the hypothesis that flight evolved prior to echolocation and shows how features associated with flight may have developed, with digits

elongating, and interdigital membranes expanding incrementally over evolutionary time. The absence of an intermediate form between a flightless proto-bat and the earliest recorded fossil bats has raised many questions over what events may have led to the evolution of flight amongst bats (Simpson, 1953; Speakman, 2001). This absence may be due to the fact that no intermediate form was preserved. This has been supported by a study that indicates that the bat fossil record is impoverished with up to 61% of its history being absent (Teeling et al., 2000). An alternative hypothesis which has been put forward is that the evolution of powered flight in bats may have been a relatively rapid evolutionary event (Sears et al., 2006) that did not result in the persistence of an intermediate form. Subsequent incremental adaptations of this structure may have then resulted in the current wing structures seen in extant bats with their variety of specialised shapes.

1.4 MOLECULAR EVENTS IN BAT LIMB DEVELOPMENT

The absence of fossil evidence demonstrating the evolution of bat wings does not preclude scientific investigation into the mechanisms behind this event. Recent studies have described how to use molecular and developmental techniques in conjunction with the fossil record to reveal how evolutionary events may have occurred (Shubin et al., 1997; Tabin et al., 1999). One strategy has been to examine commonly known limb developmental pathways, which have been highlighted through research using the chick and mouse limb. Patterns of gene expression found in these 'model' limb developmental pathways have been compared to those found in the developing bat limb. Differences that have been identified between the patterns of expression of these candidate genes have been argued to play important roles in the evolution of the bat wing.

Limb development progresses through the temporal and spatial interaction of a multitude of genes. A family of genes that play a vital role in pattern specification processes are the *Hox* genes, with *Hoxa* and *d9-10* being important for limb specific development. Late stage *Hox* expression involving *Hoxa13*, and the 5′ *Hoxd* genes (*Hoxd10*, *11*, *12* and *13*) occurs in the mouse autopods and is involved in the growth and differentiation of the digits (Montavon et al., 2008; Tarchini and Duboule, 2006). *Hoxd13* was the first gene to be characterised in bat limb development. Its expression was found to be remarkably similar both between the bat (*Carollia perspicillata*) and the mouse limbs and between bat fore- and hindlimbs. However in later stages of development patterning differences were seen, with the anterior boundary of

Hoxd13 expression in the bat hand plate becoming shifted posteriorly relative to that of the mouse (Chen et al., 2005). A more recent study confirmed these results in embryos of a different bat species (*Myotis lucifugus*), and found an additional region of expression in the leading edge of the plagiopatagia (Ray and Capecchi, 2008). They also examined the Global Control Region (GCR), an enhancer sequence known to be responsible for regulating the expression of the *Hoxd* genes (Spitz et al., 2003), finding several nucleotides that were either conserved or absent in both bat species examined (Ray and Capecchi, 2008). It was postulated that this change in expression of *Hoxd13* could play a role in the elongation of bat limbs through the interaction with two limb developmental pathways that have been previously characterised in the bat, the Bone Morphogenic Protein (BMP) and the Fibroblast Growth Factor (FGF) pathways (Ray and Capecchi, 2008).

Both an upregulation of *Bmp2* expression in the forelimbs (Sears et al., 2006) and an expansion of *Fgf8* expression in the apical ectodermal ridge (AER) (Cretekos et al., 2007) has been found in *C. perspicillata* embryos. *Fgf8* was also expressed in the interdigital tissue of the bat hand plate and it was proposed that this promoted cell survival in the presence of *Bmp* expression thereby preventing apoptosis in this region (Weatherbee et al., 2006). In addition to this, a novel pattern of *Sonic hedgehog* (*Shh*) expression has been found in the autopods of *M. natalensis* embryos (Hockman et al., 2008). *Bmp2* has also been implicated as a link between *Shh* and *Gremlin* expression, leading to a *Shh-Fgf* feedback loop in the bat autopods (Hockman et al., 2008). All of these gene and their corresponding pathways interact with and/or are regulated indirectly in some manner by the *Hoxd* gene products in the developing limb (Salsi et al., 2008; Zakany et al., 2007) making *Hoxd* genes strong candidates responsible for the unique development of the bat limb (Ray and Capecchi, 2008; Sears, 2008). Another recent publication in the field of bat limb development examined the *Paired-related homeobox* gene (*Prrx1*, also known as *Px1*), which is known to promote limb skeletal elongation (Cretekos et al., 2008). This gene was shown to have a different expression pattern between the mouse and the bat handplate. *Prrx1* was strongly expressed in the CS16 bat hand and had no expression in the mouse E12.5 hand. At CS17 the expression changed and become localised in the perichondrium of the skeletal elements. These differences were found to be driven by changes in the *cis*-regulatory regions that control *Prrx1* expression.

An alternative non-candidate gene approach would be to use microarrays to identify genes which have differences in mRNA transcript abundance between (the developing) bat hand and foot plate. In this way previously uncharacterised genes would be able to be identified as potential candidates in addition to making a list of 'the usual suspects'. This would provide a powerful and novel way of approaching the question of bat limb development and create new avenues of research in the field.

1.5 PROGRESSION OF THIS DISSERTATION

This study investigated the physical and molecular events that control the development of the bat wing. This involved the collection of embryos from the bat species, *Miniopterus natalensis*. At the same time data on the reproductive timing of this bat species was collected, and the methods of collection and the determination of pregnancy were tested (Chapter 2). Reproductive events in *M. natalensis* were further characterised by creating a developmental staging system of the embryos obtained (Chapter 3). This staging system allowed the development of *M. natalensis* to be compared with that of other bat species. The development of the fore- and hindlimbs was closely examined to determine at what stage these structures diverged and what morphological events occurred to create these unique morphologies. This information was used to determine which developmental stages were to be used for microarray analysis. A microarray experiment was performed to compare the transcriptome of the bat autopods at two different stages of development and that of the mouse at one stage of development (Chapter 4). It was assumed that genes which showed a difference in mRNA transcript abundance between bat hand and foot plates, or between bat and mouse hand plates, could play important roles in regulating the development of the bat wing. This technique allowed new candidate genes to be isolated in an unbiased manner. This information can be used for further research into the development of the unique bat wing (Chapter 5).

Chapter 2

Reproduction in *Miniopterus natalensis*

2.1 INTRODUCTION

2.1.1 *Miniopterus natalensis*: The study species

The bat species chosen for this study was *Miniopterus natalensis* (Smith, 1936), a small, gregarious, insectivorous bat. The genus *Miniopterus* (Bonaparte, 1837) is characterised by having a highly elongated third digit that folds back on itself when the wings are closed, thus they are commonly referred to as long-fingered or bent-winged bats (Taylor, 2000). *Miniopterus* has conventionally been classified as being the only member of the Miniopterinae, a sub-family belonging to the diverse family Vespertilionidae (Nowak, 1994). However, it has recently been proposed that the Miniopterinae should constitute their own family, as a sister group of the Vespertilionidae (Gopalakrishna and Chari, 1983; Miller-Butterworth et al., 2007). The *Miniopterus sp.* are widely distributed and can be found in Africa, Europe, Asia and Australia (Nowak, 1994; Simmons, 2005). *Miniopterus natalensis* was considered to be a subspecies of *M. schreibersii*; and has been referred to in the literature as *M. schreibersii natalensis*. However, based on morphological, embryological and molecular data, it is now recognised as a distinct species (Miller-Butterworth et al., 2005; Simmons, 2005).

Miniopterus natalensis has relatively small rounded ears, a high forehead, a small snout without a noseleaf or slits and their tail is long and fully enclosed in the tail membrane (Taylor, 2000). Their fur colouration ranges from slate-grey to brownish-black and russet (Taylor, 2000) and their wing membranes are dark brown (Skinner and Smithers, 1990). When fully extended the wing is long and narrow with a 'pointed swallow-like shape' (Taylor, 2000). This elongate shape means that it has a high aspect ratio (the ratio of wingspan to wing area used to describe the wing shape). This trait corresponds to aerodynamic efficiency, low manoeuvrability, and fast flight, features found commonly in open air feeders and migratory bats (Findley et al., 1972). The ability of the third digit to fold back on itself has been proposed as a novel mechanism to enhance thrust during flight (Nudds, 2007).

2.1.2 Habitat and migration of *Miniopterus natalensis*

Miniopterus natalensis is widely distributed throughout Africa. They have been recorded from eastern to southern Africa (Fig. 2.1), with their range extending up to the Sudan (Simmons, 2005).

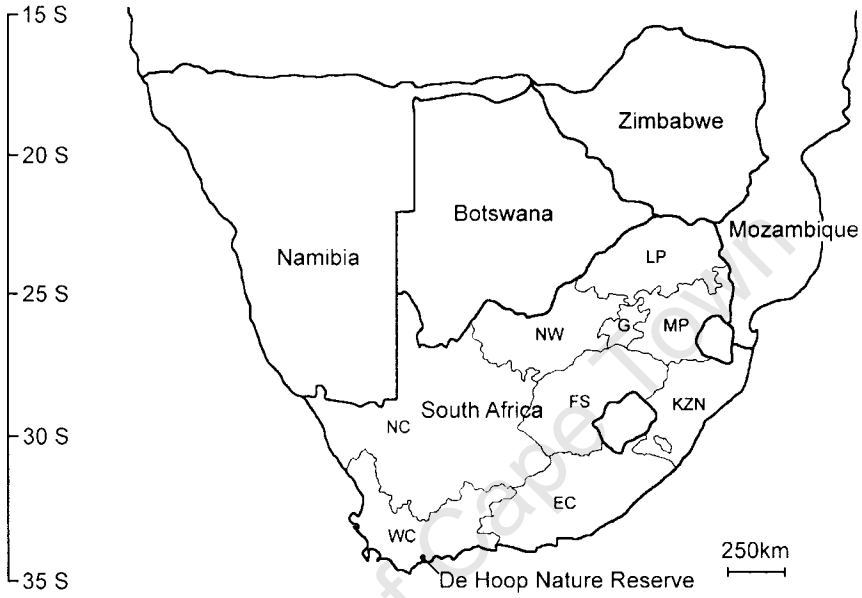


Fig. 2.1: The distribution of *Miniopterus natalensis* in southern Africa (in grey) and the study site in De Hoop Nature Reserve where the bats were sampled. Degrees of latitude are shown and the provinces of South Africa are demarcated (EC, Eastern Cape; FS, Free State; G, Gauteng; KZN, KwaZulu-Natal; LP, Limpopo Province; MP, Mpumalanga; NC, Northern Cape; NW, North-West Province; WC, Western Cape.) Adapted from Miller-Butterworth et al. (2005).

Their habitat encompasses a range of different vegetation types, from deserts and savannah to temperate Mediterranean-type scrubland, and they are commonly found in coastal regions (Jacobs et al., 2004; Miller-Butterworth et al., 2005; Taylor, 2000). Though they are able to occupy a variety of different habitats, they are limited to areas which offer suitable roosting sites. These are usually caves and abandoned mines and tunnels (Mills and Hes, 1997). The roosting needs of these bats differ seasonally. During hibernation in winter they need cool, humid roosts (known as hibernacula) that are protected from predators and the external environment (Brown and Bernard, 1994; Norton and van der Merwe, 1978; van der Merwe, 1973c). During the breeding season in summer the pregnant females need warm, humid roosts (known as maternity roosts) to give birth and establish nurseries (Rodrigues and Palmeirim, 2007). The suitability of these roosts impacts on the survival of the bats and their young (Brown and Botstein, 1999). *Miniopterus natalensis* will thus migrate long distances between suitable hibernacula and maternity roosts (van der Merwe, 1975).

As pregnant females were required for this study, it was important to review what is known about the seasonal movements of *M. natalensis* to determine when the breeding colonies would establish themselves in the maternity roosts. During the late summer and autumn (March to May in South Africa) *M. natalensis* come together in hibernacula. They begin hibernation in May and remain in torpor throughout the winter, until the end of July or early August (van der Merwe, 1973a; 1973b; 1973c). In late winter or spring (August) they leave the hibernacula and migrate to the maternity roosts which are often found in warmer areas (van der Merwe, 1973b; 1973c) . The bats tend to leave the maternity roosts at the end of summer (February and March) to migrate back to their winter roosts. In De Hoop Nature Reserve (34° 25′ S; 20° 21′ E; Fig. 2.1), Western Cape Province, South Africa, there is a breeding colony of 250 000 individuals (McDonald et al., 1990b) which occupies the De Hoop Guano Cave during the summer months (August to March). Such large breeding colonies make *M. natalensis* a suitable candidate for this study as removing small numbers of individuals should have a relatively minor impact on the whole population.

2.1.3 Reproduction in *Miniopterus*

Studies have shown that the timing of pregnancy and the gestation period of *M. natalensis* differs with latitude and environmental conditions such as temperature and rainfall (Bernard, 1980, 1994; Bernard et al., 1996; Cumming and Bernard, 1997). Copulation is thought to occur when the bats gather together in hibernacula (Bernard et al., 1996; van der Merwe, 1986), however fertilisation is determined by the initiation of ovulation (after copulation) in the female bats. *Miniopterus* uses the strategy of delayed implantation of the fertilised embryo to delay pregnancy over the winter months. Ovulation takes place in the left uterine horn for the majority of *M. natalensis* individuals and fertilisation occurs a few days after copulation and the zygote develops until it reaches the stage of a bilaminar blastocyst (van der Merwe, 1979, 1986). The fertilised zygote transmigrates to the right uterine horn which is the site of implantation (van der Merwe, 1986). During the winter hibernation period, the blastocyst floats unattached in the lumen of the right uterine horn (Bernard et al., 1996; van der Merwe, 1986). Implantation of this blastocyst occurs towards the end of hibernation. This delay ensures a period of reproductive limbo where no embryonic development occurs (Valentine et al., 1999). Post-implantation delays also occur in *Miniopterus* (Crichton et al., 1989; Kimura and Uchida, 1983; Wallace, 1978), but this has not been shown to occur in South African species (Bernard, 1994) and was not a consideration in this study. *Miniopterus*

individuals from the Eastern Cape do show a period of variation in foetal growth rate occurring in the third month of development (Bernard, 1994). This results in a period of asynchronous development amongst individuals towards the end of gestation.

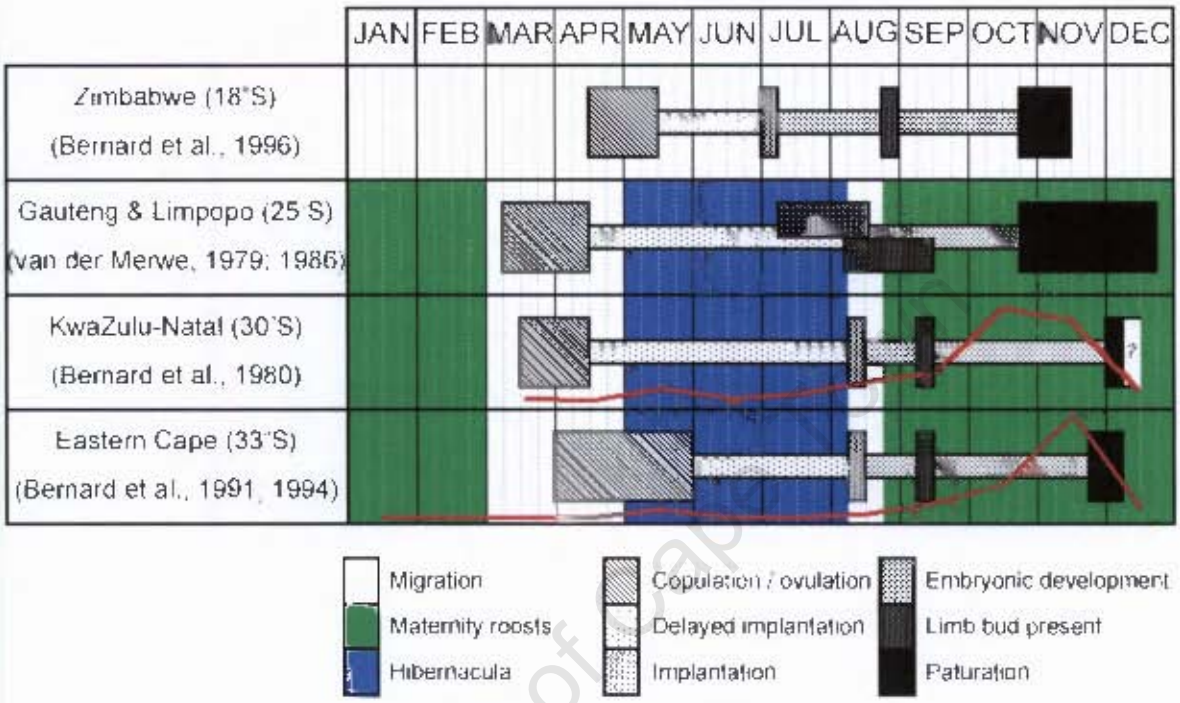


Fig. 2.2: A summary of the seasonal migration events and roosting habits of *Miniopterus* in South Africa in combination with the breeding habit and reproductive events for the different populations of this family that occur at a range of latitudes in southern Africa taken from published literature (Bernard, 1980, 1994; Bernard et al., 1991; Bernard et al., 1996; van Aarde et al., 1994; van der Merwe, 1979, 1986). Boxes indicate the range of dates over which the discrete reproduction events have been reported to occur. The red line superimposed over the data of *M. natalensis* from the KwaZulu-Natal populations and from the Eastern Cape populations shows the average monthly progesterone levels, not to scale.

Reproductive events in *M. natalensis* from a range of latitudes differ in timing and duration (Fig. 2.2). The difference in implantation dates between the tropical latitude Zimbabwean bats and the temperate latitude Eastern Cape bats seem to support the theory that implantation is governed by day length, as day length varies with latitude (Bernard et al., 1996; Racey, 1979). However, this is confounded by the catalytic activity that temperature has on reproduction (Racey, 1973). Due to the higher latitude, Zimbabwean bats are in a warmer, tropical environment while the Eastern Cape bats are in a cooler, temperate environment. This may shorten the delay of implantation in the Zimbabwean bats. Limb buds have been noted on embryos about one month after implantation for these species (van der Merwe, 1979). The phase of embryonic development takes approximately four months (Bernard,

1980; van der Merwe, 1979). The timing of birth is very similar among the different populations and is not a discrete event within the population. The maternity roost at the De Hoop Nature reserve has a similar latitude (34°S) to that of the Eastern Cape *M. natalensis* population (33°S) (Bernard, 1994; Bernard et al., 1991). The breeding season of the De Hoop population has been described to occur from October to January (McDonald et al., 1990b) with parturition occurring in individuals of the colony in late November (pers. observation, 2005). Thus it could be assumed that the reproduction and pregnancy events in the De Hoop population could follow those of the Eastern Cape population, with implantation occurring in mid-August and the limb bud stage embryos appearing around mid-September.

2.1.4 The role of progesterone in bat reproduction

Delaying implantation is a strategy that allows bats to draw out pregnancy. It enables them to mate before hibernation when they gather together in the hibernacula; carry the un-implanted and inactive blastocyst throughout the winter and give birth to young in the spring at the onset of the rainy season when insect prey should be abundant (Cumming and Bernard, 1997). Though the mechanism behind this strategy is not fully understood, it is thought to be under hormonal control. During pregnancy, progesterone plays an important role. It prepares the uterine lining for implantation and maintains this lining during the early stages of pregnancy. Thus the plasma progesterone levels of the mother bat could be used as a potential indicator of the progression of pregnancy. Maternal plasma progesterone concentrations in some bats follow a seasonal bimodal distribution (Buchanan and Younglai, 1986; Burns and Easley, 1977). A slight increase occurs after fertilisation, this has been correlated to the compaction of the zygote in the blastocyst (Pakrasi and Tiwari, 2007; van Aarde et al., 1994). Delayed implantation takes place after this event, with the blastocyst remaining inactive and unattached in the uterus (van der Merwe, 1986). An increase in progesterone levels have also been noted to occur just before implantation (Buchanan and Younglai, 1986). The progression of pregnancy and embryonic development in bats has been reported to correlate with a gradual rise in progesterone levels during early embryonic development and a rapid rise in progesterone levels during the foetal growth phase (Buchanan and Younglai, 1986; Currie et al., 1988), with a positive linear relationship being found between estimated foetal age and plasma progesterone levels (Bernard et al., 1991). Progesterone levels fall dramatically immediately after birth, in some cases dropping to about 12.5% of their original level (Currie et al., 1988).

Plasma progesterone levels of the KwaZulu-Natal and Eastern Cape populations of *M. natalensis* follow this bimodal distribution (van Aarde et al., 1994), with both populations having a very similar progesterone profile (Fig. 2.2). The characterisation of the progesterone profile of this species may offer a non-invasive indicator of timing pregnancy in bat populations. If there is a relationship between the maternal progesterone level and the development of the embryo this would provide a high resolution indicator of the progression of pregnancy, a tool that would be useful in future ecological studies.

2.1.5 Characterisation of the sampling species

The purpose of this study was to investigate the timing of reproduction and migration of *M. Natalensis* in the Guano Cave at the De Hoop Nature Reserve. It was essential to determine the commencement and progression of pregnancy in these bats, as embryonic samples at specific stages of development were required for two distinct studies that examined bat limb development. A study by Dorit Hockman, who required embryos from the early limb developmental stages for *in situ* experimentation to look at the expression of *sonic hedgehog* (*Shh*) (Hockman et al., 2008). The second was this study, in which later limb development stages were needed. When using model organisms (such as the mouse) in developmental studies, timed-matings are often performed to obtain a specific developmental stage. In this way the duration of the pregnancy can be correlated to the stage of the embryo. This has been done for a captive bred population of bats, *Carollia perspicillata,* to develop a staging system for bats (Cretekos et al., 2005). However, this could not be done for *M. natalensis* as they do not survive well in captivity. Instead, pregnant females had to be caught in the wild at times that were based on estimation of pregnancy and palpitation of the abdomen (Racey, 1969). The purpose of this research was to characterise whether pregnancy was synchronous in the *M. natalensis* population of de Hoop Nature Reserve, and to test whether embryonic stages could be resolved using maternal features such as body mass, belly palpitation and progesterone levels. This report will aid and inform future studies to enable the efficient sampling of *M. natalensis* for reproductive and developmental studies.

2.2 METHODS

2.2.1 Study site and date

Bats were sampled in the De Hoop Nature Reserve (34° 25′ S; 20° 21′ E) in the Western Cape of South Africa (Fig. 2.2). Sampling clearance was given by the University of Cape Town Faculty of Science Animal Experimentation Committee (AEC#: 2006/V4/DJ and 2008/V16/DJ) and a permit to collect the female bats from the de Hoop Nature Reserve was approved by the Western Cape Nature Conservation Board (AAA004-000030-0035). The pregnant bats were taken from the De Hoop Guano Cave, a maternity roost that is known to have a large *Miniopterus* colony present during the breeding season (McDonald et al., 1990b). A preliminary field trip was done from 18th to the 20th August 2006 to assess the condition of the *M. natalensis* population at De Hoop Nature Reserve. Two adult female bats were collected using a hand-net. Adult females with prominent nipples were chosen as reproductively active specimens (Racey, 1969). Based on the reproductive status of these two females, another fieldtrip was undertaken from the 6th to the 27th of September 2006. Additional collections were made from the 23rd September to the 5th October 2008.

2.2.2 Collection and storage of bat samples

Bats were caught using an Austbat 3-bank harp trap (Faunatech, Mount Taylor, Victoria, Australia) placed near the entrance of the cave to intercept bats at emergence. The position of the trap was changed on subsequent nights to ensure that the bats did not become habituated to its presence. The trap was erected before emergence, which usually occurred at 19h00 in September. Only female bats with large, apparent nipples were kept as specimens. These were considered to be parous females (those that had bred in previous seasons), and were chosen because they were more likely to be pregnant (Racey, 1969). Pregnant bats could also be distinguished by their relatively large bellies. Bats were transported to the field laboratory individually in soft cloth bags. These bags were transported in a secure carry-box with adequate ventilation. Bags were hung overnight and the bats were processed the following morning.

Measurements were taken to ensure that the candidate bats were *M. natalensis* and not *M. fraterculus*, a morphologically similar species that has been reported to occupy similar habitats (Stoffberg et al., 2004). The hind-foot length, from the heel to the tip of the claws

(*HL*) and the total length, from nose-tip to the end of the tail (*TL*) of the bat was measured to the nearest millimetre with vernier callipers. These measurements were put into a field identification function (Eq. 2.1) (Stoffberg et al., 2004).

$$(HL \times 0.279417) - (TL \times 0.989306) + 100 \qquad \text{(Eq. 2.1)}$$

If the result of the function was negative (< 0) the individual was considered to be *M. natalensis*, if the result was positive (> 0) the species was considered to be *M. fraterculus*.

Suitable candidates were chosen from the bats captured. Due to the limited number of permitted samples, and the importance of obtaining an exact number of specific embryo stages, great care needed to be taken when choosing the bats to be dissected. Errors in choice could result in: (i) non-pregnant bats being dissected resulting in no embryo being obtained for that bat; (ii) bats that were very early or very late in their pregnancy being dissected, resulting in embryos of an unrequired stage being obtained; or (iii) bats being dissected in which the embryo is at a stage in which the required samples had already been collected resulting in the embryo being redundant. Bats with the largest bellies (Section 2.2.3) were chosen as the candidate bats to be dissected. However as the sampling progressed, the bats were selected based on embryo stage requirements. Bats that were not selected for dissection were given water and released the following evening after capture. Selected bats were euthanized using Halothane-M&B (Safe Line Pharmaceuticals Pty. Ltd., Wadeville, RSA). Halothane is a volatile anaesthetic that can be absorbed by inhalation. It is an effective, rapid euthanizing agent suitable for use with small (<7 kg) mammals (American Veterinary Medical Association, 2001). Cotton wool was soaked in halothane and placed in a sealable container. Perforated tinfoil was placed over this cotton wool to prevent the bat from coming into direct contact with the halothane. Bats were placed in the sealed container for a minimum of 5 min. They were then removed and immediately decapitated using a pair of scissors or cervically dislocated. For some specimens, blood was collected for progesterone testing (Section 2.2.4). Bats were dissected according to procedures outlined by Behringer (pers. comm.). The length and diameter of the uterus was measured to the nearest millimetre. The Crown-Rump length (CRL) of the embryos was taken while they remained in the amniotic sac. The developmental stage of the embryo was then determined by using embryonic features outlined in the staging system of Cretekos et al. (2005). Embryos to be

used for RNA extraction were stored in RNA*later*® (QIAGEN, Valancia, CA, USA) according to the manufacturer's instructions (QIAGEN®, 2006). RNA*later*® is a storage solution that prevents the degradation of RNA in tissue samples by RNases (Gorokhova, 2005). The embryos in RNA*later*® were stored at 4°C for 24 hrs and were then stored at -20°C for the remainder of the field trip (less than a month) after which they were put in long-term storage at -80°C. The embryos used to describe the developmental stages of *M. natalensis* and those used in *in situ* experimentation (in a parallel study by D. Hockman) were stored in 4% Paraformaldehyde (PFA) (Hockman et al., 2008). This solution acts as a fixative by cross-linking proteins in the tissue sample. The embryos in 4% PFA were stored at 4°C for twenty-4 hrs; they were then dehydrated through a series of methanol washes (25%, 50%, 75% and 100% methanol). Once in 100% methanol the embryos were stored at -20°C for the remainder of the field trip (under a month). The adult bats specimens were stored in 70% ethanol; these specimens were used in additional research (Junker et al., 2008) and subsequently deposited in the mammal collection of the Transvaal Museum.

2.2.3 Examination of maternal features

Bats were weighed to the nearest 0.01 g using a Scout® *Pro* Balance (Ohaus, Pine Brook, New Jersey, USA) and their forearms measured to the nearest millimetre using vernier callipers. Belly palpitation was used to estimate embryo stage by placing the thumb and middle-finger on either side of the belly and gently squeezing or stroking it with light pressure. Though most bat bellies felt fairly turgid gentle probing usually revealed a rigidness and roundness to the belly area (the area of abdomen found below the bats ribs) when pregnant. On the basis of the resistance of the belly wall to pressure and their relative belly size, each bat was assigned a code.

2.2.4 Examination of progesterone levels

Two different methods of blood collection were used in this study: (i) During the sampling fieldtrip in which the embryos were collected, trunk blood was collected directly from the neck of the decapitated bat; (ii) During an additional fieldtrip to obtain time course data for progesterone levels, blood was collected by bleeding live bats. Once enough blood (under 200 µl) had been taken, the wound was allowed to congeal and the bats were monitored to ensure that they had stopped bleeding and given water. All of the above procedures were carried out humanely using standard procedures and guidelines (Diehl et al., 2001; Morton et al., 1993). Collected blood was spun in a bench top centrifuge at 14 000 rpm for 7 min. The

separated plasma was pipetted and put into a clean eppendorf tube and stored at -20°C for up to a month. These samples were processed by C17/GSH Chemical Pathology Laboratory, Groote Schuur Hospital. The progesterone levels were measured on the Roche MODULAR E170 automated electrochemiluminescent immunoasssay system. All data was analysed using Microsoft® Excel 2007 and Statistica v. 8.0 (StatSoft Inc., 2007).

2.3 RESULTS

2.3.1 Reproduction in *M. natalensis*

The number of bats roosting in the maternity chamber increased from a few hundred to several thousand over the period of the study (18th August to the 25th September 2006). A dramatic increase in the density of bats during emergence was noted in mid-September 2006. In mid-August 2006 the bats sampled were either not pregnant or at a very early stage of pregnancy and no embryos were obtained, however in early-September 2006 the first bat sampled was pregnant and carried an embryo at the limb bud stage. This indicated that the period of pregnancy during which the sampling needed to take place had begun. In 2006 incremental sampling sessions showed a progressive increase in the average developmental stage obtained that was weakly positively correlated ($r = 0.724$; $p < 0.01$; $n = 54$) to the date of sampling (Fig. 2.3A). In 2008, two sampling sessions held approximately two weeks apart were successful in obtaining the full range of embryonic stages required ($n = 80$; Fig. 2.3B).

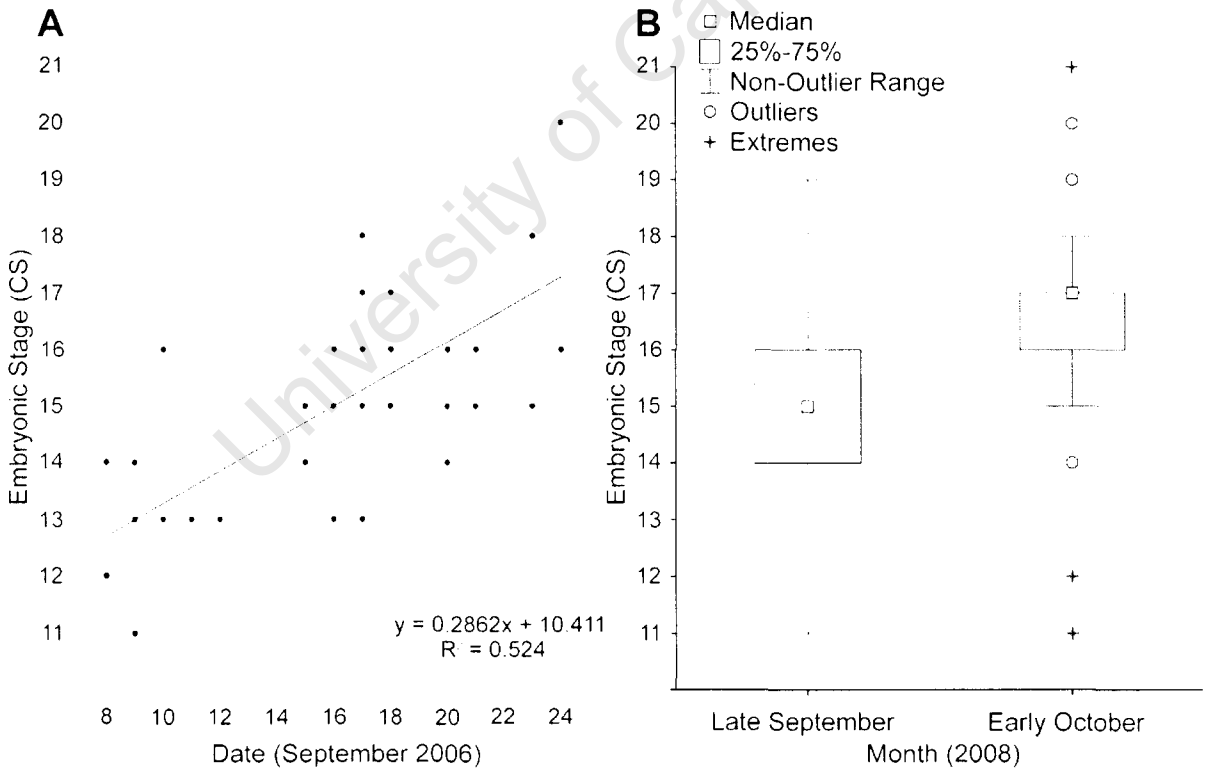


Fig. 2.3: Daily sampling sessions showed high variation was seen in the stages of the obtained embryos. (A) In 2006, incremental sampling sessions showed that embryonic stage was weakly correlated to the sampling date. (B) In 2008, two sampling sessions held approximately two weeks apart were successful in obtaining the full range of embryonic stages required, with early stages being more prevalent in late September while late stages were found in early October.

The largest range was observed in October 2008 when both CS11 and CS21 embryos were found (Fig. 2.3B). In both sampling years, there was never a day where embryos from only one stage of development were found.

2.3.2 The detection of pregnancy in *M. natalensis* in the field

The mass of the pregnant females varied greatly (Fig. 2.4) and was not correlated to the developmental stage of the embryos that they carried ($r = 0.142$; $p = 0.23$; $n = 73$). However the average mass of the pregnant bats (11.03±0.84) was significantly different from that of the non-pregnant bats (10.29±0.99; Mann-Whitney U, $z = -2.6$ $p < 0.01$; $n = 83$).

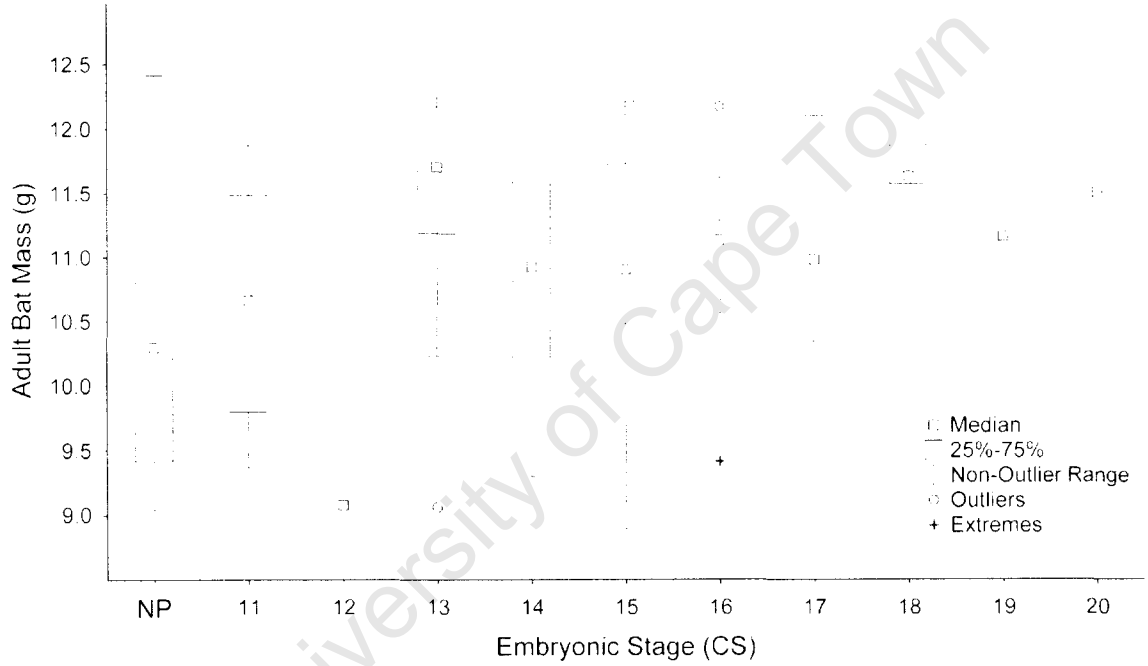


Fig. 2.4: Box plots of the mass of pregnant and non-pregnant (NP) *M. natalensis* adult bats indicate that there was no relationship between this and the stage of development of the corresponding embryo..

The belly codes provided a useful parameter by which to sort and designate bats suitable for dissection. An examination of the distribution of the embryonic stages for the progeny of the bats assigned each belly code, showed that there was a trend of earlier stages being found in bats with lower belly codes and later stages being found in bats with higher belly codes (Fig. 2.5). The positive correlation between embryonic stage and the belly code given to the maternal bat was significant but weak (Embryonic stage = $0.99x + 11.18$; $r = 0.761$; $p < 0.01$; $n = 91$). However the positive relationship between these two variables was weak. The wide range of stages found within each group of bats assigned a belly code indicated that specific stages of embryos could not be distinguished through belly palpitation.

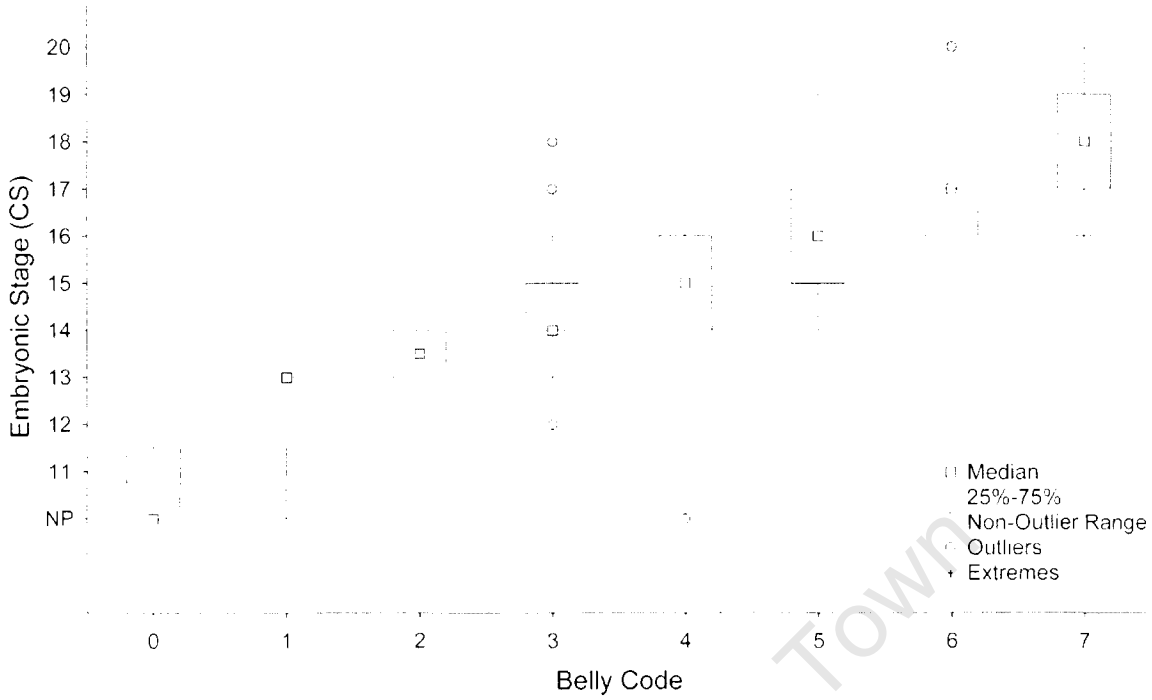


Fig. 2.5: Box plots of the distribution of the developmental stages of the embryos of pregnant *M. natalensis* adult bats that were assigned specific belly codes. The positive trend between these two variables indicate that these codes may be a useful, if low resolution, predictor when determining the progression of pregnancy.

By decreasing the resolution of the belly codes, the progression of the pregnancy (namely: non pregnancy, early development and late development) could be determined with a reasonable amount of accuracy (Table 2.1). In each of the above categories over 70% of the bats were assigned the correct belly code.

Table 2.1: Codes and description of the belly characteristics that were used to judge the relative state of pregnancy in *Miniopterus natalensis*. The number of bats that were assigned the codes is given along with the percentage that was found to be in the specified states of pregnancy.

Code	Description of bat belly	Number	Percentage (%)	State of Pregnancy
0	No belly, flaccid stomach that dips inwards	8	75	Not Pregnant
1	Small belly, sides are concave below ribs	52	71	Early Development (CS11–15)
2	Small belly, tight stomach, sides are concave below ribs			
3	Medium belly, turgid stomach, no roundness, sides are flush with ribs			
4	Medium belly, turgid rounded stomach, sides are flush with ribs			
5	Medium belly, turgid rounded stomach, sides protrude from beneath ribs			
6	Large belly, turgid stomach bulges out in front and from beneath ribs	37	81	Late Development (CS16-20)
7	Large belly, turgid rounded stomach, can see distinct protrusion on right hand side			
8	Very large belly, turgid and rounded, sides of belly protrude greatly from all regions below ribs			

2.3.3 Progesterone levels in *M. natalensis*

The plasma progesterone concentrations were, significantly lower in the first half of September 2006 (11.9±2.9 ng/ml) as compared to the second half of September 2006 (18.4±9.19 ng/ml; Mann-Whitney U z = -3.3, p = 0.0006; n = 46). The plasma progesterone concentrations obtained in September and November 2006 followed the same trend as that of the populations from KwaZulu-Natal (van Aarde et al., 1994) and the Eastern Cape (Bernard et al., 1991) with the progesterone levels increasing dramatically in November 2006 (Fig. 2.6). During November 2006 samples were taken from two heavily pregnant individuals, the progesterone levels varied greatly between these individuals, from 20.3 ng/ml to 72.5 ng/ml. Plasma progesterone levels were found to be very low in the non-reproductive and lactating individuals (7.9±1.3 ng/ml) that were sampled in that same month. Over the progression of development the maternal progesterone levels increased, becoming more variable (Fig. 2.6).

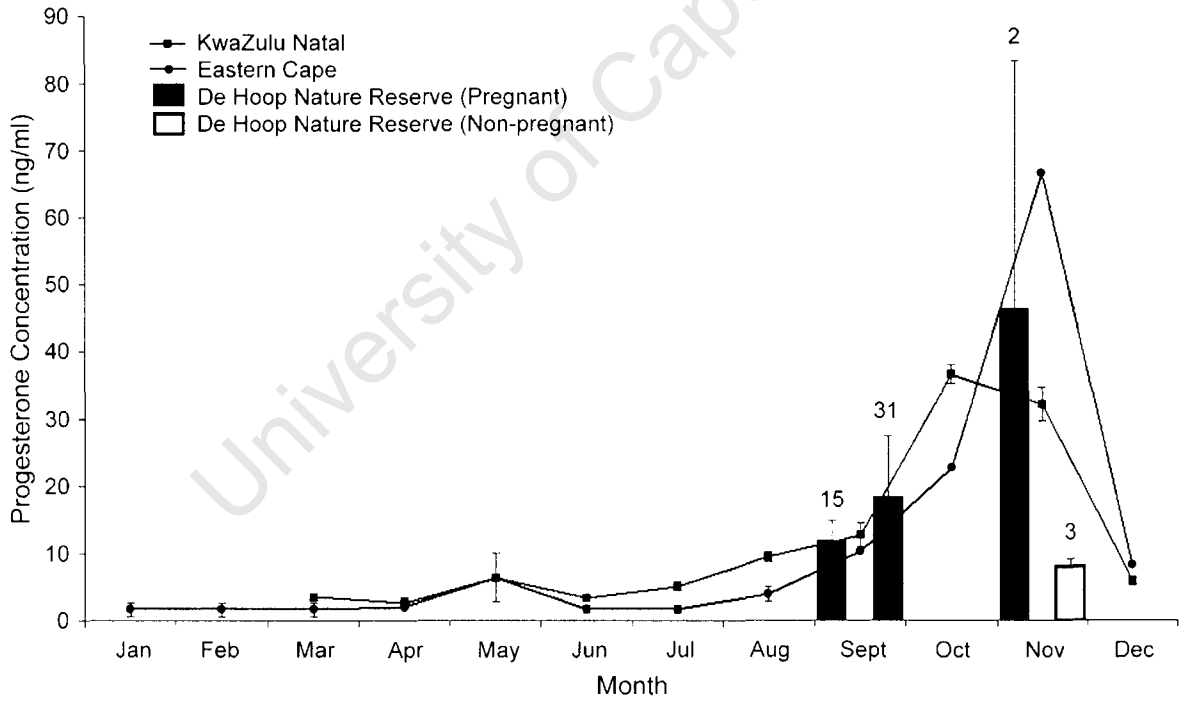


Fig. 2.6: Maternal plasma progesterone concentrations of *M. natalensis* at De Hoop Nature Reserve for September and November 2006 (bar graph) follow the same trend as the populations from KwaZulu-Natal (van Aarde et al., 1994) and the Eastern Cape (Bernard et al., 1991). The number of bats sampled in each month for this study is given above the bars and the standard deviation is shown by the error bars.

Maternal progesterone levels in bats carrying embryos in early developmental stages (CS11 to CS15) were significantly lower (13.8±4.6 ng/ml) than those found in bats carrying late developmental stages (CS16 to CS20; 29.4±10.3 ng/ml) embryos (Mann-Whitney U, z = -3.7,

$p < 0.001$ $n = 47$) (Fig. 2.7). The positive correlation between the maternal progesterone levels and embryonic stage of development was weak (Log progesterone concentration = $0.07x + 0.146$; $r = 0.652$; $p < 0.01$; $r = 0.66$; $n = 47$).

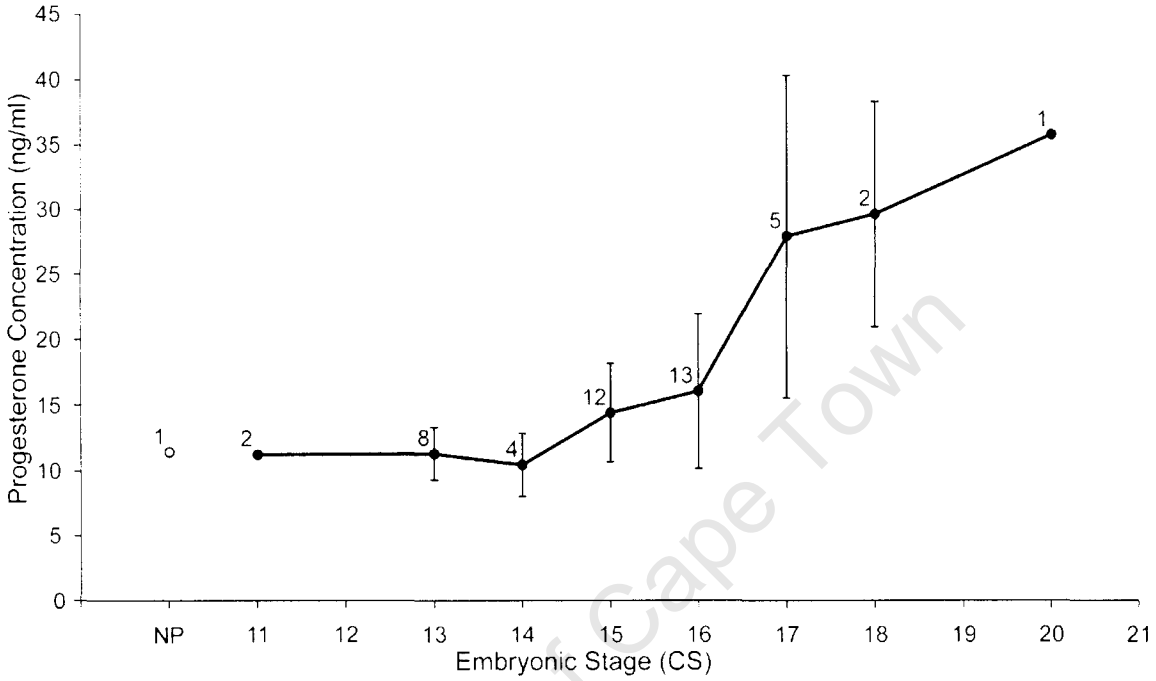


Fig. 2.7: Maternal plasma concentration progesterone in *M. natalensis* at different stages of embryonic development. The open symbol (○) represents one non-pregnant (NP) individual. (●). The sample size is given next to each value and the standard deviations are given by the error bars.

2.4 DISCUSSION

2.4.1 The timing of reproductive events is similar among populations

The timing of migration and reproduction events found in the *M. natalensis* population of the De Hoop Nature Reserve appeared to be similar to that of the populations from the KwaZulu-Natal and the Eastern Cape. The relatively small number of bats in De Hoop Guano Cave from 18th to 20th August and the fact that the two bats sampled were not pregnant, suggests that this population in the cave was made up of bats that had not migrated to a winter roost (Taylor, 2000). It is likely that, at this time, the main breeding population had not yet returned from their winter roosts. The increased density of roosting females found in the maternity chamber in September 2006 was indicative that the main part of the breeding colony of *M. natalensis* had returned to the De Hoop Guano Cave in early September. This is similar to the recorded appearance in late August of *Miniopterus* in maternity caves elsewhere in South Africa. The majority of bats caught appeared pregnant, with the samples obtained showing a progression of pregnancy within that population. The period of early to late development (CS11 to CS21) occurred over a period of about two weeks in both years sampled. This period began in early September 2006 and in late September 2008. The embryonic stages obtained were not the same among individuals sampled on the same day. Embryos ranging over ten stages of development were found on the same day indicating that though reproductive events may be synchronous (within the period of a month) in this species exact developmental events were not highly correlated between individuals. This finding indicated that incremental sampling (done daily as in 2006) is unnecessary, instead the required stages of embryos may be obtained with fewer sampling sessions (two as in 2008), that obtain larger numbers of samples.

2.4.2 Belly palpitation provides a useful measure of pregnancy

Two field based maternal features were explored as possible ways of predicting the reproductive state and stage of pregnancy in the bat: mass and belly size. Although pregnant bats were, on average, heavier than their non-pregnant counterparts, strong differences in mass could not be discerned at the level of the individual. There was no relationship between the maternal mass and the stage of the embryo being carried. The high variation in mass is attributable to fluctuations that occur in bats due to water loss and feeding behaviour (Studier et al., 1970). Due to this, the weight of the bat was not considered to be a useful indicator of

pregnancy for individual bats carrying embryos at this stage of development. Belly palpitation (Racey, 1969) was found to be effective in distinguishing between pregnant and non-pregnant individuals and between individuals carrying embryos which were from early (CS11 to CS15) and from late (CS16 to CS20) stages of development. The accuracy of this method (above 70%) determined by this study indicates that belly palpitation can provide a non-invasive, reliable measure of low resolution states of pregnancy for future bat research. The belly codes given to the pregnant bats were found to be weakly correlated to the stage of development of the embryo being carried. Though individual stages of the embryos could not be distinguished, the belly palpitation method provided the best method by which to sort and select bats for dissection. This technique was useful in categorising the relative stages of pregnancy among a sample group of bats. However, it is not useful in obtaining an embryo of an exact stage. It is recommended that there should be a margin of error incorporated into sampling sessions to ensure that the required embryo stages be obtained. Belly palpitation was affected by the size of the bat, if it was echolocating or struggling and by whether it had ingested anything prior to being sampled. It is recommended that this method be used under standardised conditions, when the bat is not stressed and after it has defecated.

2.4.3 Progesterone levels are highly variable within pregnancy

The maternal plasma progesterone concentrations were found to differ between early and late September. They rose in November and dropped immediately after birth (in lactating individuals that were sampled at the same date) following the pattern described by (Bernard et al., 1991; van Aarde et al., 1994). The weak correlation found between progesterone concentrations and the embryonic stage was due to the high variation found in late stage embryos. This variation may have been a result of the low sample numbers within these stages. Low resolution differences in progesterone concentration (between early and late stage embryos) however, were not a reliable indicator of embryonic stage. Progesterone levels have been shown to increase in a step-wise fashion with crown-rump length in the little brown bat (Currie et al., 1988), with the greatest increases occurring during the phase of foetal growth (the period of growth post-organogenesis). However these stages were not sampled in this study. Due to its high variability maternal plasma progesterone concentration would most likely be a more useful indicator to determine larger scale events during pregnancy (such as discrimination between non-pregnancy, early embryonic development and late foetal growth) of a population rather than small scale (embryonic stage) events of an individual.

Chapter 3

Bat embryonic development: Breaking the mould of model organisms

3.1 INTRODUCTION

3.1.1 Bat developmental staging systems

There has been a longstanding curiosity about the morphological development of bat embryos (Spillmann, 1925). However, despite this, the field of bat embryonic development has remained a relatively small one, with reproductive studies typically looking at implantation and very early stages of embryonic development (Gopalakrishna and Karim, 1979; Rasweiler IV, 1979; Richardson, 1977; van der Merwe, 1996; Wallace, 1978). There have been several reports describing the post-natal development of bats and their wings (Isaac and Marimuthu, 1996; Jones, 1967; Reiter, 2004). Detailed descriptions of late-stage embryos and the progression of foetal development in bats are lacking, leaving a gap in our knowledge of bat development. More recently, Lawrence (1991) produced a biological study of *Syconycteris australis* from New Guinea, which included a description of foetal development. However, the presentation of the foetal development in this bat was limited by the lack of previous research in this field and the need for more embryological studies in this order was highlighted. Development of seven stages in *Myotis lucifugus* embryonic development has also been briefly described by Adams (1992a).

Recently a comprehensive embryonic staging system based on timed matings for the short-tailed fruit bat (*Carollia perspicillata*) was published (Cretekos et al., 2005). This study provided the first comprehensive staging system that can be used for studying bat development. It highlighted the differences that occur in the developmental progression of the fore- and hindlimbs of the bat (Fig. 3.1). Based on morphology, the bat forelimb development was found to deviate from that of the mouse from the formation of the hand plate onwards. Asymmetry was visible earlier, the digits elongated and the interdigital tissue did not regress. In contrast to the bat hand plate, the bat foot plate did not show as great a morphological difference to that of the mouse. However its digits were remarkably symmetrical (Fig. 3.1)

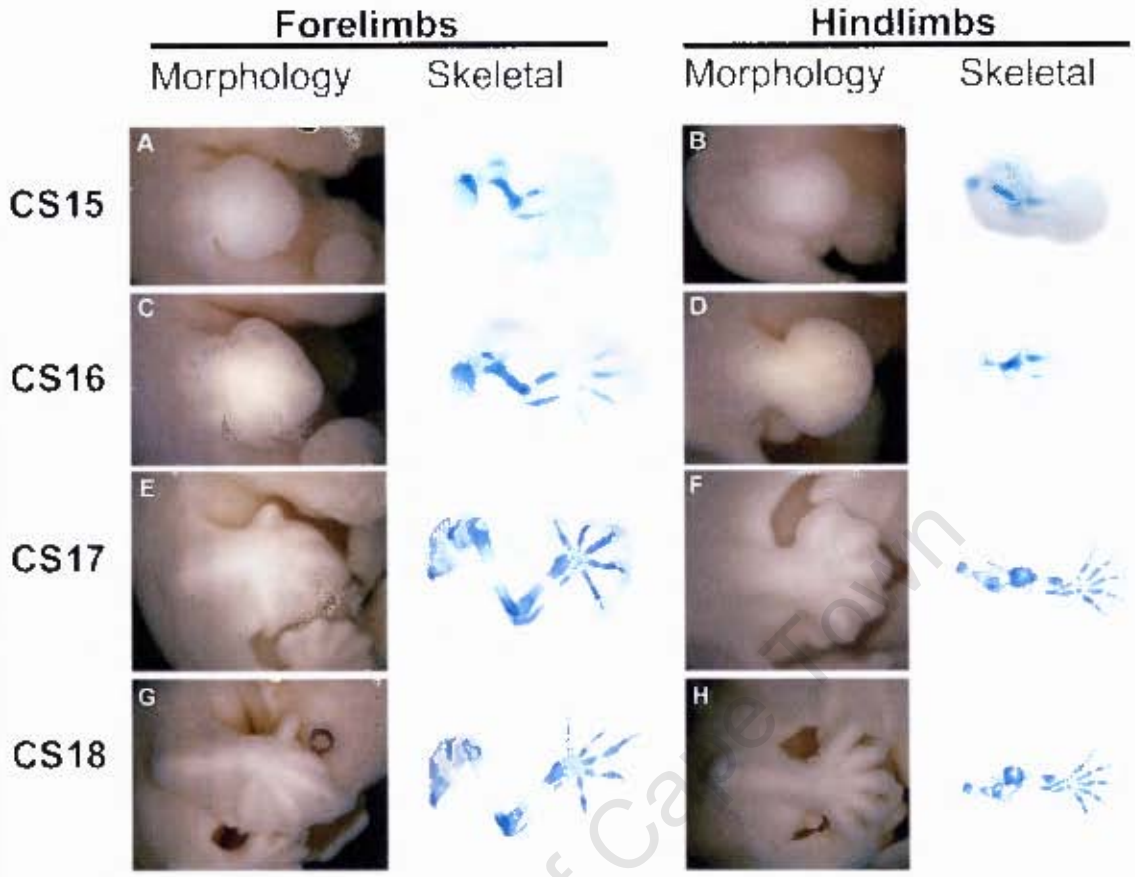


Fig. 3.1: Fore- and hindlimbs of *C. perspicillata* from stages 15 through to 18. The morphology of the limb is shown on the left while equivalently staged limbs on the right have been stained for cartilage (blue). Centres of ossification can be seen as a white area within the element (Cretekos et al., 2005)

The Cretekos et al.(2005) staging system was found to be a useful guide for staging two other phyllostomid bats, namely *Desmodus rotundus* and *Phyllostomus hastatus*. However species-specific differences in stages were found even within this closely related group. The functional application of the *Carollia* staging system was subsequently used to describe development in other bat species (Fig. 3.2). The embryonic development of Japanese house bat, *Pipistrellus abramus* was examined and compared to that of *C. perspicillata*. The crown-rump length (CRL) was consistently smaller throughout development in *P. abramus* and species-specific differences were found occur in late development. Further characterisation of adult-related species-specific differences found in developing bats was done in a study embryonic development in the Black mastiff bat, *Molossus rufus* (Nolte et al., 2009). Wing development, in particular, was shown to have features that related to the adult morphology indicating that their shape and structure are defined in early development (Nolte et al., 2009). Both *P. abramus* and *M. rufus* share a close phylogenetic relationship to *M. natalensis* (Fig. 3.2) (Eick et al., 2005). The development of the integumentary structures was examined in

Geoffroy's rousette, *Rousettus amplexicaudatus* (Giannini et al., 2006). This species is more phylogenetically distant from *C. perspicillata* than either *P. abramus* or *M. natalensis* (Fig. 3.2) (Eick et al., 2005). Once again embryos at later stages of development tended to deviate more from the *Carollia* staging system, and refinements in staging terminology was proposed (Giannini et al., 2006). The *Carollia* staging system was found to be a useful framework by which to stage embryos of all the above species, with these additional studies contributing towards the understanding of particular species-specific differences that occur during embryonic development (Giannini et al., 2006; Nolte et al., 2009; Tokita, 2006).

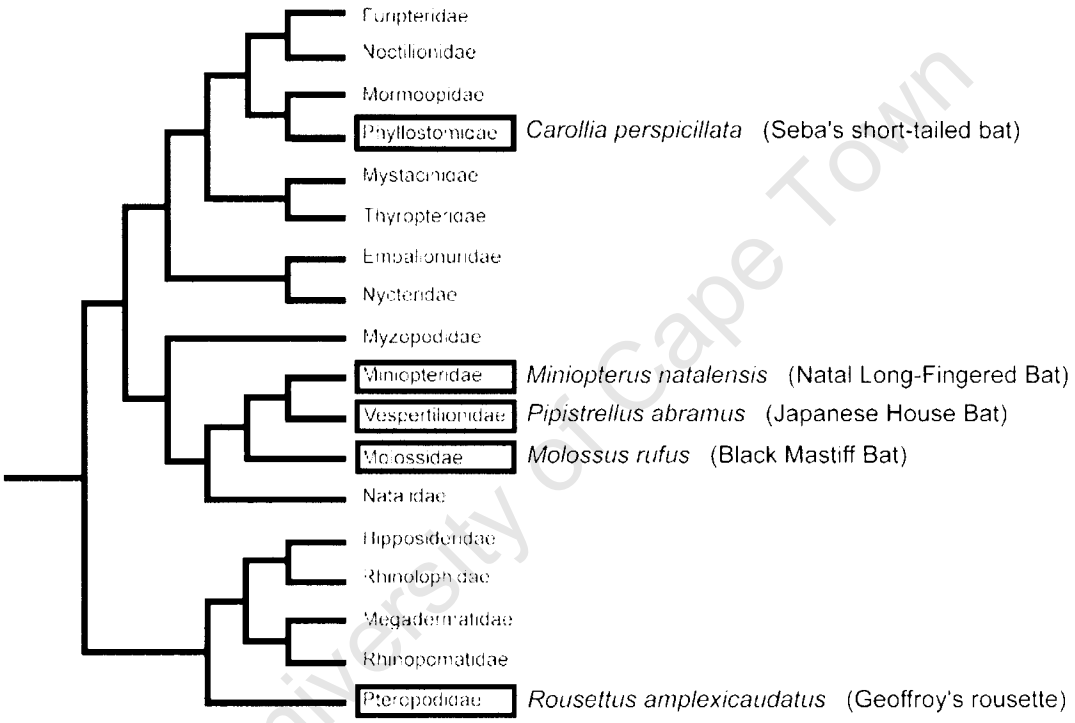


Fig. 3.2: Phylogenetic relationships between the different families of bats. *Miniopterus natalensis*, *Pipistrellus abramus* and *Molossus rufus* share a closer familial relationship to each other than to *Carollia perspicillata*. The familial relationship of *Rousettus amplexicaudatus* is shown to be the most phylogenetically distant from *C. perspicillata*. The phylogeny and figure has been adapted from Eick et al. (2005).

The creation of the *Carollia* staging system for bat development has laid the foundation for future molecular research. Several molecular studies have used *C. perspicillata* in research looking at bat wing development (Chen et al., 2005; Cretekos et al., 2007; Sears et al., 2006; Weatherbee et al., 2006). The purpose of this study was to characterise the development of an African bat species, *M. natalensis*, to lay the foundation for subsequent molecular developmental studies on this species. The progression of development was noted and the development of the fore- and hind limbs was described.

3.2 METHODS

3.2.1 Embryonic samples

The embryos were fixed overnight at 4°C in 4% paraformaldehyde (PFA) (Sigma, St. Louis, MO, USA) in PBS. Following fixation, the embryos were dehydrated through a methanol series for storage at -20°C in 100% methanol. Embryos used for skeletal analysis were fixed in 10% buffered formalin, they were transferred to 95% ethanol and stained for cartilage and calcium (bone) using Alcian blue and Alizarin red (McLeod, 1980).

3.2.2 Photography

Photographs were taken of the embryos at magnifications ranging 1 X to 9 X. This was done using a Nikon Stereoscopic Zoom Microscope SMZ1500 equipped with a Nikon Digital Sight Camera Control Unit (DS-U2) and a DS-5M Camera head (Nikon Instruments Inc., Melville. NY, USA) and NIS-Elements (Nikon Instruments Inc., Melville. NY, USA) digital 3D imaging software. The embryos remained submerged in ethanol during this process. Photographs were edited using Adobe® Photoshop® CS2 (v. 9.0). *Miniopterus natalensis* embryos were staged according to the *Carollia* staging system (Cretekos et al., 2005), embryonic development was examined and compared to that of *C. perspicillata*, *P. abramus* (Tokita, 2006), *R. amplexicaudatus* (Giannini et al., 2006) and *Mus musculus* (Kaufman, 1992; Wanek et al., 1989). Original stage names given by Cretekos et al. (2005) were used, except for CS16 and CS19 where the stage nomenclature of Giannini et al. (2006) were used due to species-specific differences, such as the absence of a nose-leaf in *M. natalensis*. A new stage name is given for CS21, which was not described in Cretekos et al. (2005). Data were analysed using Microsoft® Excel 2007.

3.3 RESULTS

3.3.1 Progression of development in *M. natalensis*

A total of 127 *M. natalensis* embryos were collected, ranging in stage from CS12 to 21 (Table 3.1). The uterus and the crown-rump length of *M. natalensis* embryos increased in a linear fashion, indicating that there was a relatively constant rate of growth over these stages (Table 3.1; Fig. 3.3). This rate of growth was comparable to that seen in *P. abramus* (Tokita, 2006) but lower than that seen in *C. perspicillata* and *M. rufus* (Cretekos et al., 2005; Nolte et al., 2009).

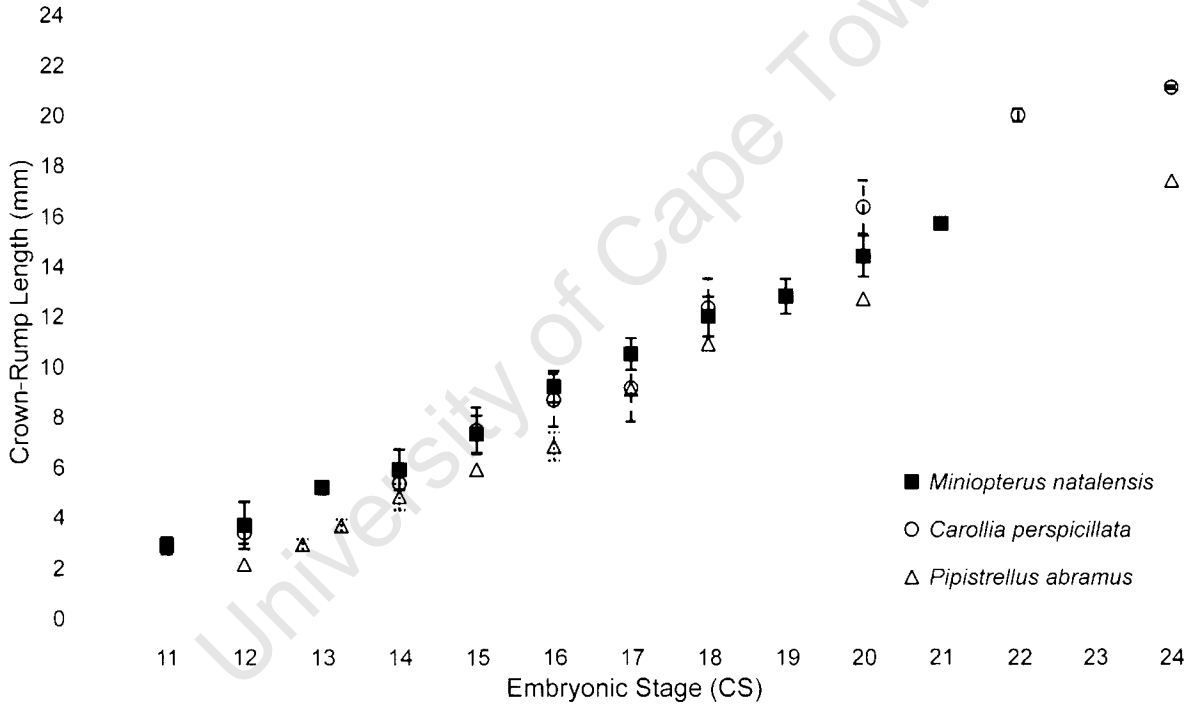


Fig. 3.3: Changes in the mean embryonic crown-rump length during development for *M. natalensis* (squares), *C. perspicillata* (circles; Cretekos et al., 2005) and *P. abramus* (triangles; Tokita, 2006). Error bars indicate standard deviations from the mean when three or more measurements were available.

A summary of the stages, key developmental events and characteristics particular to *M. natalensis* development are highlighted in Table 3.1. Embryos could be easily staged using the *Carollia* staging system (Cretekos et al., 2005) along with the major features of mouse development (Kaufman, 1992). Details of embryonic development, specifically pertaining to that of the limb morphology, at these stages follows.

Table 3.1: Staging and measurements for *M. natalensis* embryos including the key limb features used to determine the stage based on Cretekos et al. (2005) and the number of specimens obtained for each stage. The average uterus and crown-rump measurements are shown in bold (±standard deviation) with the range of measurements given below.

Stage (CS)	Key limb bud features	No. embryos	Uterus width (mm)	Uterus length (mm)	Crown-rump length (mm)
12	**Forelimb (FI)**: Slight protrusion adjacent to somites 7-11.	3	**4.6** (±0.19) 4.5-4.8	**5.8** (±0.12) 5.7-5.8	**3.7** (±0.94) 3.0-4.3
13	**FI**: Limb bud forms a distinct semi-circular bulge. **Hindlimb (HI)**: Slight protrusion adjacent to somites 23-27.	11	**5.4** (±0.53) 4.5-6.0	**8.0** (±2.09) 5.8-10.0	**5.2** (±0.26) 5.0-5.5
13L	**FI**: AER present. **HI**: Forms a semi-circular bulge.				
14E	**FI**: Flexure divides the proximal from the distal limb bud; limb bud is as long as it is wide. **HI**: AER present.	18	**6.0** (±0.71) 4.3-7.0	**9.0** (±1.01) 7.8-10.5	**5.9** (±0.80) 5.0-8.0
14	**FI**: Limb bud is longer than it is wide and lies against the flank of the embryo; propatagium primordium present as bulge on proximal-anterior edge. **HI**: Limb bud is as long as it is wide.				
14L	**FI**: Distal FI is wider along anterior-posterior (A-P) axis than proximal FI; plagiopatagium primordium (Plp) present as bulge on proximal-posterior edge. **HI**: Anterior side longer than posterior side forming curved asymmetrical bud.				
15VE	**FI**: Project outwards at right angles to the body axis; symmetrical hand plates (HP) present. **HI**: Distal HI is wider along A-P axis than proximal HI but still asymmetrical.	21	**7.4** (±0.78) 5.7-8.5	**10.4** (±0.81) 8.7-12.0	**7.3** (±0.75) 6.2-9.0
15E	**FI**: Thumb primordium protrudes slightly from anterior Hp. **HI**: Symmetrical foot plate (FP) present.				
15	**FI**: Distinct thumb protrusion; Plp bulges outwards; digit condensations visible. **HI**: Symmetrical FP with smooth edge.				
16 E	**FI**: Edge of interdigital tissue between digits 4 and 5 takes on slight convex shape. **HI**: FP wider at base than distally due to proximal expansion; digit condensations visible.	26	**8.2** (±0.81) 7.0-10.0	**11.8** (±0.83) 10.7-13.5	**9.1** (±0.68) 7.8-10.0
16	**FI**: Digit primordia extend beyond HP edge. **HI**: Interdigital tissue begins to recede.				
17	**FI**: Interdigital tissue between the thumb and digit 2 recedes. **HI**: Tips of FP digits are free due to apoptosis of interdigital tissue.	31	**9.1** (±0.51) 7.7-10.3	**12.2** (±1.02) 11.0-15.8	**10.5** (±0.63) 9.0-11.8
18	**FI**: Thumb completely free. **HI**: Interdigital tissue receded half way along the digits	10	**10.0** (±0.68) 8.8-10.8	**13.0** (±1.33) 11.8-16.0	**12.0** (±0.80) 10.8-13.5
19	**FI**: Claw primordia develop on thumb. **HI**: Claw primordia develop; digits are completely free of interdigital tissue.	4	**10.4** (±0.52) 8.8-10.8	**13.1** (±1.28) 11.8-14.5	**12.8** (±0.69) 10.8-13.5
20	**FI**: Digits curve inwards; thumb claws partially keratinised. **HI**: Claws partially keratinised.	2	**10.1** (±0.12) 10.0-10.2	**14.3** (±0.35) 14.0-14.5	**14.4** (±0.82) 13.8-15.0
21	**FI , HL**: Claws are fully keratinised and appear iridescent; chiropatagia expand and become translucent	1	**10.0**	**15.3**	**15.7**

3.3.2 CS12: Forelimb bud

Embryos at this stage are characterised by the appearance of the forelimb bud (Fig. 3.4C & D). These appear as a dorso-ventrally (D-V) flattened protrusion extending from the lateral ridge. They lie adjacent to the seventh through to the eleventh somite in an embryo with 27 somite pairs. Cranial flexure is seen (Fig. 3.4A) and three pairs of pharyngeal arches (mandibular, hyoid and glossopharyngeal) extend distally from each side of the cranial region with the mandibular arch becoming more bulbous towards its distal end (Fig. 3.4A). The anterior two arches, which appear in CS11, are larger. The length of the forelimb bud base is much wider than the proximal-distal (P-D) length. At this point the buds are slightly asymmetrical, with the maximum P-D length being found in the posterior portion of the buds. No hindlimb buds are present (Fig. 3.4B). Optic evaginations, which form in CS11, occur near the lateral forebrain (Fig. 3.4A & B). The otic pits which form in CS11 have closed to form spherical otic vesicles which enlarge to form clearly visible protuberances on either side of the caudal part of the hindbrain vesicle (Fig. 3.4A).

3.3.3 CS13: Hindlimb bud

Embryos at this stage are characterised by the presence of the hindlimb buds, which first appear as a bulge arising from the caudal region of the lateral ridge (Fig. 3.4F). They lie adjacent to the twenty-third through to the twenty-seventh somite in an embryo with 32 somite pairs. They are similar in appearance (shape and symmetry) to the CS12 forelimb bud, though they are smaller in size. The hindlimb bud grows during this stage to form an asymmetrical mound of tissue. The forelimb bud (Fig. 3.4F & G) has increased in size along the P-D axis. It appears as a distinct, symmetrical mound of tissue, with the sides of the bud sloping evenly both anteriorly and posteriorly. The apical ectodermal ridge (AER) is visible as a prominent line that runs along the A-P axis of the distal tip of the bud (Fig. 3.4H). The mandibular arch is larger and is now composed of a bulbous distal mandible (lower jaw) and a proximal maxilla (upper jaw and palate) (Fig. 3.4E). These are separated by a cleft called the oral groove which becomes narrower as the two structures enlarge (Fig. 3.4E). The distal portion of the glossopharyngeal arch has become concealed behind the hyoid arch, as this stage progresses it becomes increasingly obscured. The optic vesicle has become depressed within the enlarging cephalic region. However it remains clearly evaginated (Fig. 3.4E & F). The formation and elongation of the endolymphatic appendage can be seen as a dorsal protuberance from the otic vesicle (Fig. 3.4E & G). Late in this stage (in a 35 somite embryo)

an indentation corresponding to the nasal pit can be seen on the ventral surface of the nasal process, dividing it into a lateral and a medial nasal process. The naso-lacrimonial groove is also apparent. Very late in this stage the forelimb bud can be seen to bend ventrally and a slight bulge begins to develop on the anterior-ventral surface of the proximal limb. The forelimb bud becomes dome-shaped with its A-P axis being equivalent to its D-V axis.

3.3.4 CS14: Pigmented retina

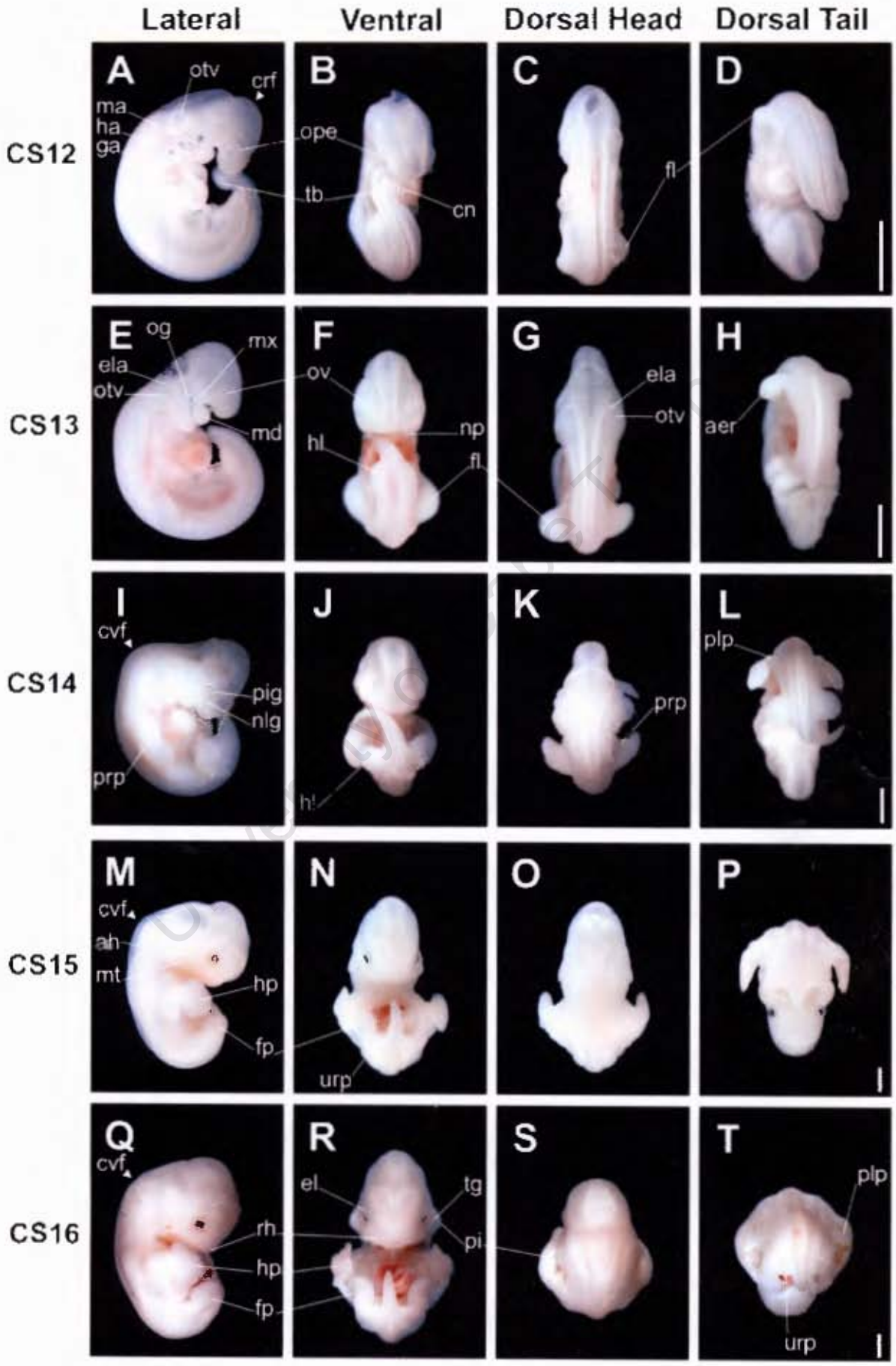
The embryo has undergone cervical flexure creating a bend of approximately 90° between the head and the trunk (Fig. 3.4I). The body axis progressively straightens during this stage while the tail narrows and elongates. The forelimb bud (Fig. 3.5A) is longer along its P-D axis than its anterior-posterior (A-P) axis, P-D elongation increases as development within this stage progresses. The protuberance of the AER is still clearly visible along the distal edge of the bud. The forelimb buds have become increasingly bent causing their P-D axis to 'point' towards the posterior. It is dome-shaped and there is no P-D regionalisation. A distinct marginal vein can be seen just beneath the AER (Fig. 3.5A). A slight bulge of tissue can be seen on its anterior and posterior portions, these are the primordia of the protopatagium and plagiopatagium respectively (Fig. 3.4K, L & Fig. **3.5**A). The hindlimb bud (Fig. 3.4J & Fig. **3.5**B) becomes increasingly longer than its width over this stage. They are not symmetrical about their P-D axis, the anterior side of the bud slopes more towards the body wall than the posterior side. The oral groove has continued to deepen as the mandiblar and maxillary components of the mandibular arch become more prominent (bulbous). The hyoid arch has enlarged further completely obscuring the glossopharyngeal and fourth arches. At this stage pigment in the retina of the embryos becomes visible in the anterior boundary of the retina. As development within this stage progresses there is an increase in pigmentation across the retina (Fig. 3.4I).

3.3.5 CS15: Hand/foot plate

Embryos at this stage were characterised by the formation of the hand and foot plates. In the CS15E embryo the distal portion of the forelimb bud (Fig. 3.5C) becomes flattened and paddle-shaped; it then proceeds to extend slightly along its A-P axis to form a hand plate. The forelimb itself has increased in its P-D length and bends at the putative elbow region. The proto- and plagiopatagia primordial swellings have become larger and more distinct (Fig. 3.5C). Towards the end of the early portion of this stage the posterior half of the hand plate becomes slightly larger than the anterior. During the transition between the early and the late

stage the hand plate begins to appear slightly warped as mesenchymal condensations begin to form. In the CS15E embryos, the hindlimb bud (Fig. 3.5D) lengthens becoming angled towards the posterior. No distinct regionalisation between the proximal and distal portion is evident. The AER is present and the marginal vein can be seen. The tissue between the posterior portion of the hindlimb bud and the body wall, the uropatagium primordium, has become more distinct. As the stage progresses the distal portion of the footplate becomes increasingly flattened and paddle-shaped. In CS15 embryos, the hand plate has a distinct blunt, 'spade-like' shape (Fig. 3.4M & Fig. **3.5**E). The anterior "putative thumb" area extends outwards sharply just distal to the area of wrist constriction. The posterior hand plate area extends in a more rounded fashion. The edge of the hand plate takes on a slightly bumpy appearance just distal to the "putative thumb" area, corresponding to the recession of this edge (arrow in Fig. 3.5E). Chondrogenesis is initiated in the hand plate at this stage and cartilage condensations can be seen in digits two to five (data not shown). The proto-and plagiopatagia have extended into prominent swellings of D-V flattened tissue anterior and posterior to the forelimb respectively. In the CS15 embryos the hindlimb (Fig. 3.5F) has a distinct footplate which is rounded and symmetrical in shape (Fig. 3.4M & N). Digit condensations form in the foot plate late in this stage and chondrogenesis can be seen in the region of digit four at this time (data not shown). At this later stage the uropatagium primordia have expanded in the area between the proximal hindlimb and the tail (Fig. 3.4N &Fig. **3.5**F). During this stage the lower jaw is formed by the merging at the midline of the distal mandibular components of the mandibular arch. The elongated groove between the mandibular and hyoid arch forms the meatus (the external auditory canal) (Fig. 3.4M). During this stage auditory hillocks (which will become the external ear) form on either side of the meatus (Fig. 3.4M), these develop from the surface of the pharyngeal arches.

Fig. 3.4: Developmental stages CS12 to CS16 (as indicated in the left-hand column) of the bat *M. natalensis* The lateral view of the embryo (with dorsal to the left) is given by the first column (A,E,I,M,Q,U); the ventral view (with anterior to the top) is given by the second column (B,F,J,N,R,V); the dorsal view of the head and trunk are given in the second column (C,G,K,O,S,W) and the dorsal view of the trunk and tail are given in the fourth column (D,H,L,P,T,X). **aer**, apical ectodermal ridge; **ah**, auditory hillocks; **chp**, chiropatagium; **cn**, caudal neuropore; **crf**, cranial flexure; **cvf**, cervical flexure; **dc**, digital condensation; **el**, eyelid primordia; **ela**; endolymphatic appendage; **fl**, forelimb bud; **fp**, foot plate; **ga**, glossopharyngeal arch; **ha**, hyoid arch; **hl**, hindlimb bud; **hp**, hand plate; **ma**, mandibular arch; **md**, mandible; **mt**, meatus; **mx**, maxilla; **nlg**, naso-lacrimonial groove; **np**, nasal pit; **og**, oral groove; **ope**, optic evagination; optic vesicle; **otv**, otic vesicle; **pi**, pinna; **pig**, pigment; **plp**, plagiopatagium; **prp**, propatagium primordium; **rh**, rhinarium; **tb**, tail bud; **tg**, tragus; **urp**, uropatagium; **vb**, vibrissae. Scale bars = 1 mm in D (applies to A–D), in H (applies to E–H), in L (applies to I–L), in P (applies to M–P), in T (applies to Q–T))


Lateral
Ventral
Dorsal Head
Dorsal Tail
CS12
CS13
CS14
CS15
CS16
A
otv
crf
ma
ha
ga
B
ope
tb
cn
C
fl
D
E
og
mx
ela
otv
md
F
ov
np
hl
fl
G
ela
otv
H
aer
I
cvf
pig
nlg
prp
J
hl
K
prp
L
plp
M
cvf
ah
mt
hp
fp
N
urp
O
P
Q
cvf
rh
hp
fp
R
el
tg
pi
S
T
plp
urp

3.3.6 CS16: Pinna primordium

Continual growth of the forelimb causes its P-D axis to point more ventrally. The hand plate is slightly curled towards the ventral surface of the embryo. Mesenchymal condensations are evident as raised or thickened areas of tissue in the areas to become the digit rays with the third and fourth digit being the first to prominently appear. Cartilage condensations corresponding to the digits have become well defined with that of digit one and five being shorter than the others (data not shown). Later in this stage a putative joint can be seen as an area of clear cartilage two thirds of the way down the length of digit three and four (data not shown). The area of bumpy ridges on the anterior portion of the hand plate is still evident. As development continues the mesenchymal condensations of the forth digit ray push out at the anterior edge of the hand plate. This is a distinctive feature of CS16L (not shown). The plagiopatagium extends across from the forelimb to the hindlimb as a flattened sheet of tissue (Fig. 3.4T &Fig. **3.5**G). It does not extend to the tip of the fifth digit nor does it extend to the base of the foot plate as reported in *C. perspicillata* (Cretekos et al., 2005). At CS16 the foot plate remains fairly symmetrical (Fig. 3.4Q, R & Fig. **3.5**H). Mesenchymal condensations are not apparent and there is no warping or distension of the foot plate. The marginal vein is still evident. As development progresses towards CS16L mesenchymal condensations begin to form. During CS16L in these embryos (not shown) these digit rays become more pronounced, they splay out evenly across the symmetrical foot plate. At this stage no ossification is evident in either the fore- or the hindlimb elements (data not shown). The uropatagium primordium begins to extend from posterior edge of the proximal hindlimb. The mandibular and hyoid arches have become indistinct as they merged to form the entire facial complex. The auditory hillocks have grown to form pinnae and tragii progenitors (Fig. 3.4R & S). Raised areas of tissue, known as the eyelid primordia, forms around the eye (Fig. 3.4R). The anterior portion of the each nasal pit is lined by a slight swelling (rhinarium, pilca alaris) and the groove between them (rhinarium, sulcus medianus) is visible (Fig. 3.4Q). The upper and lower jaw extends creating a snout-like appearance.

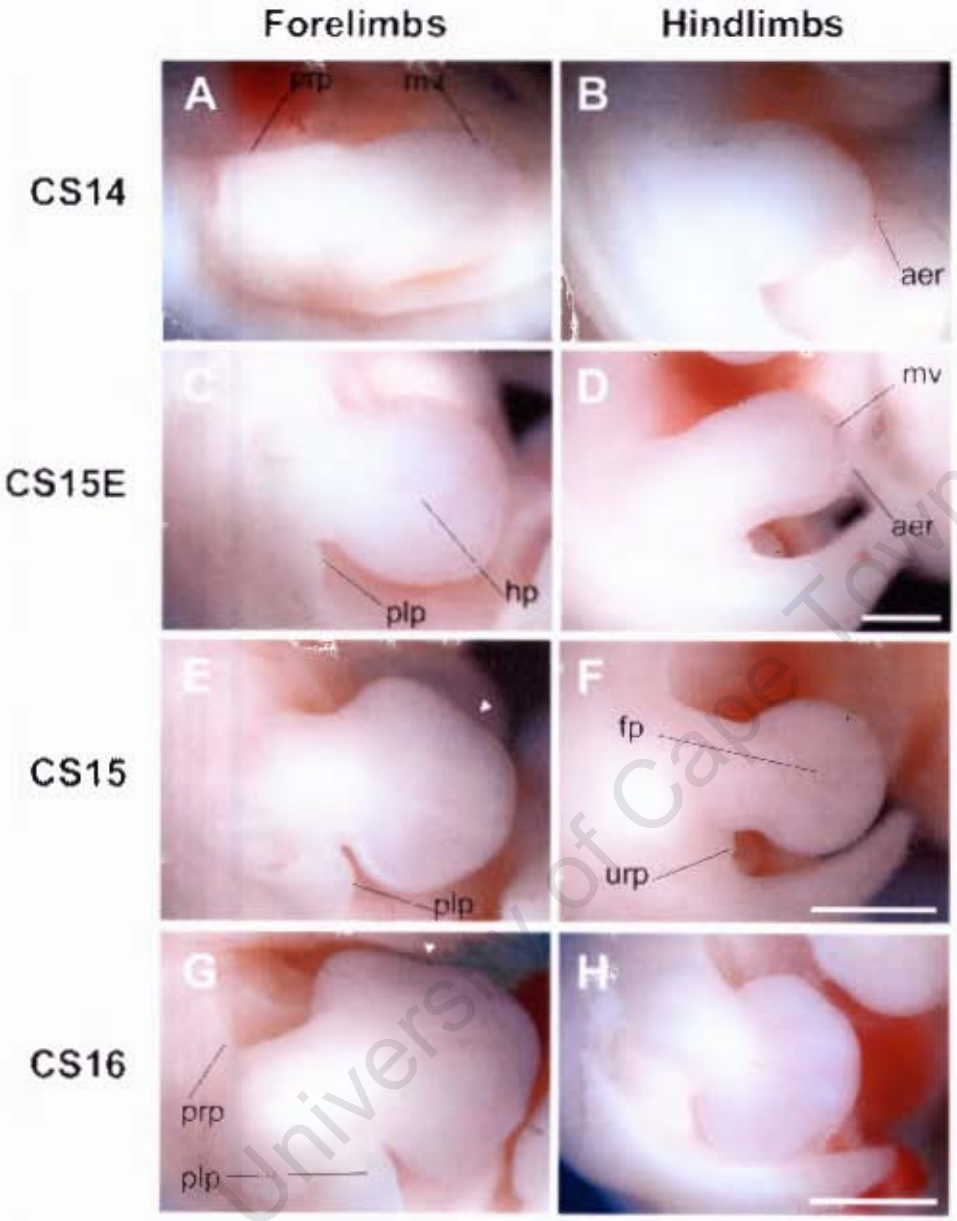
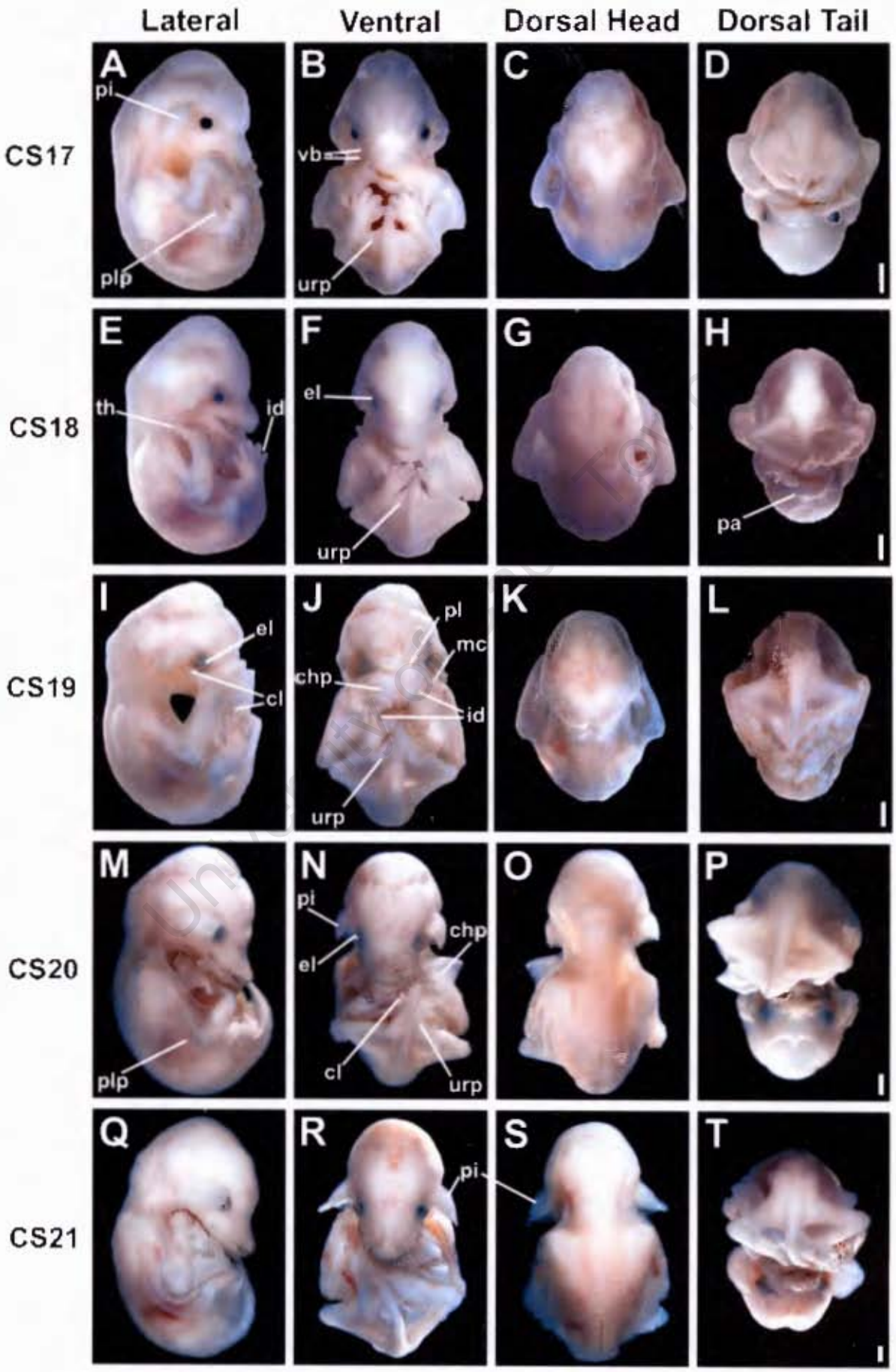


Fig. 3.5: Development of the forelimb and hindlimb of *M. natalensis* at stages CS14 to CS16. The dorsal view of the limb (with its proximal portion to the left where possible) is given. **aer**, apical ectodermal ridge; **fp**, foot plate; **hp**, hand plate; **mv**, marginal vein; **plp**, plagiopatagium; **prp**, propatagium primordium; **urp**, uropatagium. Scale bars = 0.5 mm in B (applies to A–B), in D (applies to C–D). Scale bar = 1 mm in F (applies to E-F), in H (applies to G-H).

3.3.7 CS17: Tongue out

In this stage the forelimb has become much longer and the three limb regions: the stylopodium, zeugopodium and autopodium, can be distinguished. There is slight flexure between these regions at the elbow and wrist joints. The hand plates are noticeably larger and are tucked beneath the lower jaw and the hindlimbs. The hand plate has achieved a distinct 'mitten' shape (Fig. 3.7A). The interdigital mesenchyme between the thumb and second digit has begun to recede, with the thumb seen as a protrusion approximately 90° to the P-D axis of the forelimb (Fig. 3.7A). The digit condensations are very distinctive and protrude at the edge of the hand plate and the interdigital tissue persists to form the chiropatagium (Fig. 3.7A). Joints form in the cartilage elements of the foot plate with two joints forming in digits two to five and one joint forming it the digit one (Fig. 3.7D). Late in this stage calcification is initiated in the centre of the humerus and the femur, indicating that ossification has begun in these elements (Fig. 3.7B & D). The plagiopatagia extend between the wrists and the ankles (Fig. 3.6A). It has a projection of tissue that extends beneath the wrist; it remains distinct from the hand plate at this stage (Fig. 3.7A). The foot plate undergoes a dramatic transformation due to the regression of the interdigital mesenchyme, the tips of the digit rays protrude and the more proximal interdigital tissue becomes thinner (Fig. 3.7C). The foot plate remains fairly symmetrical and the digits are equivalent in length. The knees are pronounced within the uropatagium and are able to flex slightly. The uropatagium has extended to enclose roughly half the length of the tail (Fig. 3.6B). The cervical flexure of the embryo has been reduced causing the head to flex upwards. The prominent muzzle extends distally and is matched in length by the extension of the lower jaw. The tongue does not protrude out of the mouth which is held open. Small bumps known as vibrissae, or whisker follicles, can be seen on the lateral surface of the upper jaw (Fig. 3.6B). The pinnae are rounded in shape, the dorsal tragus project sharply outwards (Fig. 3.6A). Scattered pigmentation in seen over the cheek, pinnae and dorsal neck areas.

Fig. 3.6: Developmental stages CS17 to CS21 (as indicated in the left-hand column) of the bat *M. natalensis*. The lateral view of the embryo (with dorsal to the left) is given by the first column (A,E,I,M,Q,U); the ventral view (with anterior to the top) is given by the second column (B,F,J,N,R,V); the dorsal view of the head and trunk are given in the second column (C,G,K,O,S,W) and the dorsal view of the trunk and tail are given in the fourth column (D,H,L,P,T,X). **chp**, chiropatagia; **cl**, claw; **el**, eyelid; **id**, interdigital tissue; **mc**, metacarpal; **pi**, pinna; **pl**, phalange; **plp**, plagiopatagium; **th**, thumb; **urp**, uropatagium; **vb**, vibrissae. Scale bars = 1 mm in D (applies to A-D), in H (applies to E-H), in L (applies to I-L), in P (applies to M-P), in T (applies to Q-T). Scale bars = 1 mm in D (applies to A–D), in H (applies to E–H), in L (applies to I–L), in P (applies to M–P), in T (applies to Q–T).


Lateral
Ventral
Dorsal Head
Dorsal Tail
CS17
CS18
CS19
CS20
CS21
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
pi
plp
vb
urp
th
id
el
pa
cl
pl
mc
chp

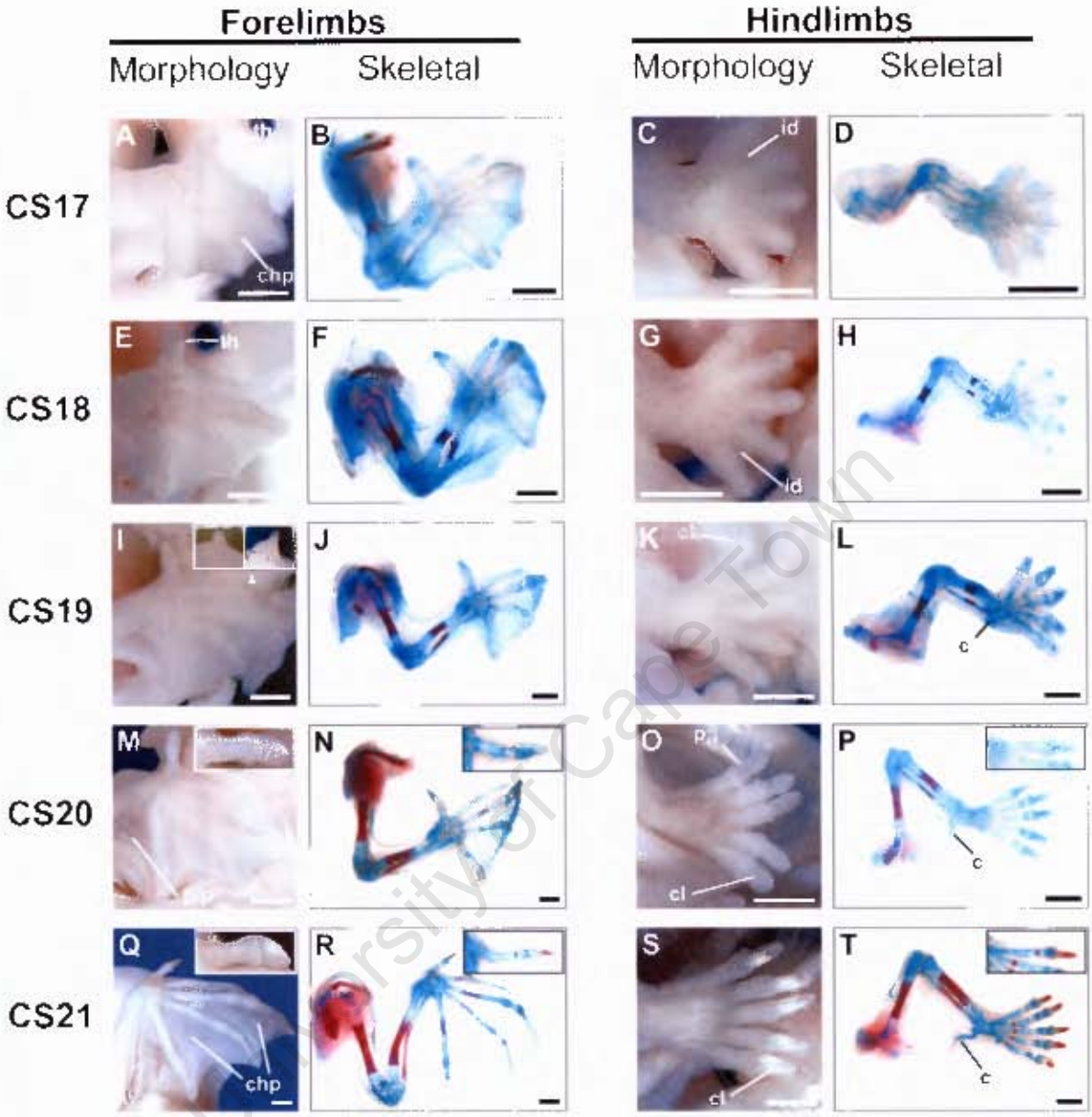


Fig. 3.7: The progression of limb and skeletal development in *M. natalensis* embryonic limbs (CS17 to CS21). Morphology of forelimbs: A, E, I, M and Q and hindlimbs: C, G, K, O and S. Skeletal staining using Alcian blue for cartilage (blue) and Alizarin red for bone (red) of the forelimbs: B, F, J, N and R and the hindlimbs: D, H, L, P and T. A magnified view of the distal tip of digit two both laterally (left) and ventrally (right) is shown inset (I). Magnified views of the thumb (M, N, Q and R) and the first digit of the foot (P and T) are shown inset. All panels show the dorsal surface of the limb with anterior toward top and proximal at left. **c**, calcar; **chp**, chiropatagium; **cl**, claw; **id**, interdigital tissue. **pg**, proximal groove; **plp**, plagiopatagium; **tm**, thumb. Scale – 1mm in all figures, inset figures are not to scale.

3.3.8 CS18: Free thumb

The embryo becomes increasingly pigmented over this stage with the areas of greatest pigmentation occurring around the lips and the eyes. The eyelids begin to extend over the eye but do not close during this stage (Fig. 3.6F). The thumb becomes bulbous towards its distal tip; it is almost completely free with only a thin section of webbing joining its base to the second digit (Fig. 3.7E). The chiropatagium is noticeably thinner in comparison to the digit rays, which form prominent ridges (Fig. 3.7E). Digits two to five of the hand plates extend (Fig. 3.7E). The digits of the foot are increasingly separated as the interdigital tissue regresses to half way along their length (Fig. 3.7G). Digit one is slightly shorter than the other digits of the foot (Fig. 3.7G & H). Ossification begins in the centre of the radius and ulna very early in this stage with ossification occurring in the tibia and the fibula later in this stage (Fig. 3.7F & H). The uropatagium extends three quarters along the length of the tail (Fig. 3.6F).

3.3.9 CS19: Claw primordium

During this stage the eyelids proceed to cover the retina (Fig. 3.6I). The entire embryo is lightly pigmented with higher concentrations of melanocytes occurring in the snout and eye regions. Digits two to five of the hands have elongated and curl inwards at their distal joints to cover the snout (Fig. 3.6J). The metacarpals have extended relative to the rest of the phalanges (Fig. 3.7I & J). The chiropatagium has expanded and thinned (Fig. 3.7I). A prominent projection of tissue extends perpendicularly (rostrally) from the tip of the second digit of the developing wing. This tissue is not a primordium of any structure. It appears to be a projection of the patagial covering (tissue that extends across the tips of digits two to five) that has not fully regressed.(Fig. 3.7I inset) The thumb is completely free and the digits of the foot are fully separated from one another (Fig. 3.7I & K). The claw primordia can be seen as distinct regions of tissue in the thumb and the digits of the foot (Fig. 3.7I & K). The stylopodal and zeugopodal elements of both the fore- and hindlimb increase in length and the areas of ossification expand (Fig. 3.7J & L). Ossification is not yet apparent in the autopodal elements. The calcar becomes apparent as an area of chondrogenesis just below the heel (Fig. 3.7L). The uropatagium has extended becoming thin and translucent; however the tip of the tail remains free of this membrane (Fig. 3.6J).

3.3.10 CS20: Eyelids closed

The eyelids remain closed during this stage. The pinnae are down-turned and have elongated triangular tips (Fig. 3.6N). The patagial extension on the tip of the second digit of the developing wing is still present though reduced in size. The plagiopatagium has extended to join the tip of the digit five to the ankle of the foot. This membrane appears translucent with opaque ridges of thickened tissue running in parallel between the digit five and the zeugopod (Fig. 3.7M). Early in this stage, a single centre of ossification is initiated in the centre of the metacarpal of digit three (data not shown). Later, the metacarpals of digits two to four are clearly ossified while a light centre of ossification can be seen in the metacarpal of digit five. The thumb and the all of the digits of the foot remain cartilaginous (Fig. 3.7N & P). A deep proximal groove divides the claw from the rest of the digit in both the thumb and the digits of the foot (Fig. 3.7M & O). The claws are pear shaped with pointed tips (Fig. 3.7M & O). Keratinisation begins to occur in the tips of the claws (Fig. 3.7M & O). The tail is fully enclosed within the triangular uropatagia and the calcar increases in length (Fig. 3.6N; Fig. 3.7P).

3.3.11 CS21: Translucent wing

The embryo is similar in appearance to the previous stage; however it is much larger in size and differentiation is seen in various integumentary structures making this stage distinct. The eyelids remain closed (Fig. 3.6R). The pinnae have expanded and flare out from the head exposing the paddle-shaped tragus (Fig. 3.6R & S). The chiropatagia, plagiopatagia and uropatagia have expanded and lie in folds against the body (Fig. 3.6Q & R). These membranes have thinned to such an extent that they are translucent with opaque ridges occurring on their surface. The limb elements have all increased in length, with digit three becoming longer than the ulna and radius, which in turn are longer than the humerus (Fig. 3.7R). Ossification has occurred over approximately half the length of the humerus and a quarter of the length of the radius and ulna (Fig. 3.7R). The regions of ossification in the metacarpals of digits two to three increase and ossification is now clear in the metacarpal of digit five (Fig. 3.7R). Ossification is still absent from the metacarpal of the thumb (Fig. 3.7R inset). It is now clear that digits one to five of the hand have the following number of phalanges respectively: 2, 1, 2, 2 and 2 (Fig. 3.7R) as in the adult (see Fig. 1.1). No ossification is seen in the phalanges of digits two to five but a thin band of ossification can be seen in both the proximal and distal phalanx of the thumb (Fig. 3.7R inset). The digits of the

foot begin ossification in the metatarsals of digits two to five, the proximal phalange of digit one and the distal phalanges of all digits (Fig. 3.7T). The claws of the thumb and the foot are iridescent indicating that they are fully keratinised. They are sharply pointed with grey areas of pigmentation beginning to appear at their tips (Fig. 3.7Q & S).

3.4 DISCUSSION

3.4.1 Development across bat species is conserved

The *Carollia* staging system was found to be an excellent template to use in staging *M. natalensis* embryos. Though these two species are from fairly distant families (Fig. 3.2) and have divergent features (e.g. a nose leaf in *C. perspicillata* and a snout in *M. natalensis*), the use of robust embryological features enabled equivalent stages to be found easily. The constant rate of size (CRL) increase found in *M. natalensis* was similar to that found in *P. abramus* (Tokita, 2006), but slightly lower that that found in *C. perspicillata* (Cretekos et al., 2005) embryos. This relationship between these species may be a reflection of their phylogenetic relationship, with *P. abramus* and *M. natalensis* being more closely related to one another than to *C. perspicillata* (Fig. 3.2). Though *M. natalensis* was similar in mass to *C. perspicillata* in early development in late development *C perspicillata* was heavier and *P. abramus* was lighter. This relationship may remain constant throughout foetal growth as reflected by the newborn weights of the three species (*C. perspicillata*: 4.7-6.4 g, (Rasweiler and Badwaik, 1997); *M. natalensis*: 2.2-3.1 g (van der Merwe, 1979); *P. abramus*: 2-3 g (Tokita, 2006)). The adult bat size and weight (head body length, HB; forearm length, FA; mass, M) also had a similar relationship (*C perspicillata*: HB = 48-65 mm, FA = 34-45 mm, M = 10-20 g (Nowak, 1994); *M. natalensis*: HBL = 55-68 mm, FA = 42-48 mm, 9.4-13.0 g (Stoffberg et al., 2004); *P abramus*: HB = 41-60 mm, FA = 30-37 mm (Tokita, 2006), M = 3.9-6.5 g (Zhang et al., 2008)). This supports the suggestion by Tokita (2006) that species-specific differences in adult size may be evident during embryonic development. This relationship may be attributable to limitations in the ability of pregnant bats to carry the additional weight of the embryo during flight (a constraint of wing-loading). This may put selective pressure on pregnant bats in terms of speed, manoeuvrability and the energetic costs of flight which would constrain the size of developing embryos (Hayssen and Kunz, 1996; Hughes and Rayner, 1991). This is an interesting feature of development in this order as it may offer the only example of foetal maternal constraint posed by flight factors. However, further work should be done on embryonic size and adult features among a variety of bat species, including additional measurements for *M. natalensis* embryos at later stages of development and of all species during foetal growth, to provide further support for this trend.

In agreement with Cretekos et al. (2005), the relative position of the limb bud initiation among bat species (*M. natalensis, P abramus, C. perspicillata*) is conserved. Few features differ in terms of timing and position between these bats for the stages described however species-specific differences are evident. In *C. perspicillata* the nose leaf primordium is apparent while *M. natalensis, P. abramus* and *M. rufus* develop a snout with a prominent rhinarium. The snout shape of *M. natalensis* appeared distinct from that of *M. rufus* from CS17 onwards. That of *M. rufus* was broad, short and flattened as compared to the rounded, beak-like snout of *M. natalensis*. This may be related to different feeding behaviours or oral echolocation abilities between these two bats (Nolte et al., 2009). The ears appear elongated and triangular in *P. abramus* in comparison to the down-turned, rounded ears of late stage *M. natalensis*. The pinnae do not migrate to the mid-line of the head in either of these two species as reported in *M. rufus*. In these species the morphological difference in pinnae development are evident by CS20. Tail development also differs among species. *Carollia perspicillata* has a distinctive, reduced tail. During development it becomes encased by the uropatagium by CS18 (Cretekos et al., 2005). The tail of *M. natalensis* and *P abramus* is elongated. During development the tail is only completely encased by CS20. *Molossus rufus* has an elongated tail that is enclosed by uropatagium halfway along its length. During development, this state is reached by CS20, with additional elongation of the tail and expansion of the uropatagium occurring in later stages. The pattern of tail development in these species mirrors their phylogenetic relationship, with the closely related miniopterids and vespertilionids showing very similar development and adult features. Developmental processes become increasingly divergent as the phylogenetic distance increases among species. The differences in the adult tail morphology between phyllostomids on the one hand and miniopterids and vespertilionids on the other are linked to differences in flight and feeding behaviour. In the latter two families the tail is longer because the tail membrane is used to catch insects (Altringham, 1996) and to provide lift during flight (Norberg and Rayner, 1987). This is not the case in the plant visiting phyllostomids.

The wings of the adult *C. perspicillata* and *M. natalensis* are very similar in shape, with both species having the same aspect ratio (*C. perspicillata* = 6.1 (Norberg and Rayner, 1987) *M. natalensis* = 6.1 (Jacobs, 1999)). In spite of this, these two bats have very different flight capabilities. *M. natalensis* is an aerial hawker of insects (and migratory; (van der Merwe, 1973c)), capable of flying rapidly in open habitats and slowly in closed (i.e. vegetated)

habitats (Jacobs, 1999; McDonald et al., 1990a). *Carollia perspicillata* has the ability of highly manoeuvrable, slow flight that enables it to hover while feeding on fruit and flowers (Stockwell, 2001). These differences may arise from slight modifications of the functional structure of the wing. One such modification is the loss of the third phalange in the third digit of the wing of *M. natalensis* when compared to *C. perspicillata*, and the relative extension of the second phalange of this same digit (Nowak, 1999). This difference becomes evident soon after joint formation at CS17 in the developing embryos. In *M. natalensis,* this change results in the maintenance of the wing length with the loss of a degree of freedom in that digit. This elongated phalange can be folded against the wing and it has been proposed that it provides a novel mechanism for this narrow-winged bat to produce thrust during slow flight within caves (Nudds, 2007) and closed habitats (Jacobs, 1999). Therefore, the plasticity of the skeletal (finger and tail bones) and integumentary (patagia and pinnae) structures during late development among different bat species, allow features to develop that can be directly related their morphology and function in the adult bat.

3.4.2 Divergence of fore- and hindlimb morphology occurs after CS15

The fore and hind limb of the bat is considered to be relatively synchronous in terms of the timing of developmental events (Bininda-Emonds et al., 2007). The developmental lag of the hindlimb was evident during early development (CS12-CS15), but this was not dramatic. The events governing the growth and morphology of the fore- and hindlimb appeared to be similar across stages CS12 to CS15. During CS15 the hand plate changed shape and became increasingly asymmetrical; however the foot plate retained its symmetrical, rounded shape and just increased in overall size. This indicates that the molecular events that determine the morphology of the hand plates are only seen shaping the hand plate at this stage of development. The 'bumpy' appearance to the anterior edge of the hand plate observed from CS15 to CS17 coincides with the regression of this area of tissue and the freeing of the thumb by CS18. The interdigital tissue of digits two to five does not regress; instead it becomes thinner to eventually form an expansive interdigital webbing (chiropatagia). There is an apparent loss of apoptosis in this tissue, paradoxically, genes that are involved in promoting apoptosis are still expressed in these areas (Weatherbee et al., 2006). It is currently unknown whether apoptotic events are still occurring in these areas. It has been assumed that there is a complete loss of the apoptosis in the interdigital regions of the bat hand plate, however if this were so, a morphology of fused digits with thick interdigital mesenchymal tissue would be expected (Emmanouil-Nikoloussi et al., 2008). Apoptosis may still be prevalent in the

interdigital tissue of the hand plate, playing a role in thinning the interdigital mesenchyme, with novel regulation preventing the complete regression of this interdigital tissue. This may be combined with proliferation of the interdigital ectoderm to create the extensive webbing in this region. In contrast, regression of interdigital tissue of all digits of the hindlimb is found to occur from CS17 in a manner similar to that found in the mouse autopods (Wanek et al., 1989). Differential elongation of the digits also appear to occur during this stage of development, with the digits of the hand plate appearing longer and non uniform in size, while those of the foot plate appear similar in size.

The size of the digits condensations appear to be larger in the hand plate than in the foot plate from CS17 onwards. This is in contrast to what was reported by Sears et al. (2006) who only found significant differences between these elements from CS20 onwards. Accurate measurements should be made on the cartilage contestations of several biological replicates before further claims can be made. The timing of cartilage condensation and the formation of centres of ossification appears to be slightly delayed in the hindlimb as opposed to the forelimb giving some indication of the prevalence of developmental lag through to CS21. However the sequence of formation of cartilage and ossification in the elements of both appendages was the same. The progression of ossification in *M. natalensis* was similar to that described in other bat species (Adams, 1992b).

At embryonic stages CS16 and CS17 it is expected that molecular events that control apoptosis will differ between the hand and foot plates of the developing bat. The divergent morphologies of these two structures also indicate that molecular events controlling cell proliferation and cell shape and growth may also be different between these structures. Molecular events that pattern the autopods and so regulate future elongation processes may also be specified, however no events related to ossification should be found. It is differences in these molecular events that determine the differences in morphology between the hand and foot plate of the bat. The events that control the changing morphology of the bat limb can only be examined on a molecular scale. This can be done using a 'bottom up' approach, where a candidate gene that has been shown to be involved in similar events as the process of interest (i.e. regulation of apoptosis) is isolated and characterised in the context of bat limb development. Or it can be done using a 'top-down' approach, where a genome-wide screening of all genes that are expressed during this limb developmental process is done

through microarray experiments. Examination of the molecular processes occurring in stages CS16 and CS17 should provide valuable information regarding the formation of the bat autopod structures. The bat hindlimb is a unique structure; however it is more comparable in morphology to the mouse limbs and can therefore be used as a point of comparison for bat wing development. Comparisons between the two structures would allow specific differences in the regulation of development to be isolated.

Chapter 4

Gene expression in the developing bat wing

4.1 INTRODUCTION

4.1.1 Microarray studies in limb development

Microarray analysis allows the simultaneous measurement of the relative expression of thousands of genes in one experiment and can be used to identify differentially expressed genes independently of any knowledge of their function. In microarray technology the transcriptome (the population of RNA molecules) of a tissue is isolated and compared to that of another tissue. It provides the ideal first step for comparative molecular development studies, allowing a low resolution, high gene through-put examination of all the gene expression occurring in the tissue of interest. This can be followed by the detailed characterisation of temporal and spatial patterns of expression of genes identified as being differentially expressed. Microarray technology has already been used to identify genes which are differentially expressed in the hand and foot plates of E12.5 mouse embryos (Table 4.1) (Shou et al., 2005). Many of these genes (i.e.: *Pitx1*, *Tbx4*, *Tbx5* and *Hoxc10*) had been identified previously in a SAGE analysis of the fore- and hindlimb E11.5 mouse embryo (Table 4.1) (Margulies et al., 2001), while others (i.e.: *Rdh10*, *Frzb*, *Tbx18* and *Hip*) were identified as novel candidates. The differences in gene expression between wild-type and mutant limbs, have also been examined using a microarray experiment on E11 mouse limbs with retinoic acid (RA) induced phocomelia (shortened long bones and limbs), identifying 111 genes with altered expression (Table 4.1) (Qin et al., 2002). A second microarray analysis study performed on mutated mouse limbs compared the mRNA (transcript) profile of wild-type E12.5 mouse hand plates and genitalia to those that lacked a functional *Hoxd* gene cluster (Table 4.1) (Cobb and Duboule, 2005). Fourteen genes were identified as potential HOXD targets, of these, *Hoxa11*, *Sgk*, *Gfra2*, *Epha3*, *Odz4* and *Gdf10* were found to be strong candidates with the latter five being regulated in the same manner in both the hand plates and the genitals (Cobb and Duboule, 2005). Gene expression differences were found to be relatively low-level in most of these experiments, with under fifty genes exhibiting fold changes higher than 2 (Table 4.1). This finding suggests that commonly expressed genes may be used in different ways to specify limb type and that these processes may be regulated by small numbers of differentially expressed genes (Shou et al., 2005).

Table 4.1: Summary of microarray studies on mouse limb development with a description of the experimental platform used, the comparisons made, the number of differentially expressed (DE) genes obtained, the fold change (FC) and p value (p) cut-offs used to obtain these lists.

Embryonic Sample	Description	Comparison	Reference	DE	FC	p
E11.5 Mouse	SAGE Analysis: 36 300 unique SAGE tags	Forelimb vs. Hindlimb	(Margulies et al., 2001)	70 14	- >2	<0.01 <0.01
E12.5 Mouse	Affymatrix (Mouse 430 A and B Chips): 26 179 probe sets detected	Pooled hand plates vs. Pooled foot plates	(Shou et al., 2005)	44 13	>2 >2	<0.05 <0.01
E11.0 Mouse	Mouse GEM™ 1 Microarray: 9 000 IMAGE Consortium clones	Mutant forelimb (RA for 6 hr) vs. WT forelimb	(Qin et al., 2002)	111	>1.5	-
E12.5 Mouse	Affymatrix (U74Av2 GeneChip Microarray): 12 000 probes	Mutant hand plate (*Hoxd*$^{Del1-13}$) vs. WT hand plates	(Cobb and Duboule, 2005; Shou et al., 2005)	12 1	>1.4 >2	<0.0025 <0.00025

In this chapter candidate genes, which may play a role in the evolution and development of the bat wing, were identified by a microarray analysis of relative mRNA transcript abundance between (i) bat hand and foot plates and (ii) bat and mouse hand plates

4.1.2 Overview of a microarray experiment

Microarrays work on the principals of nucleic acid hybridisation (Fig. 4.1), RNA from the tissue (the target) is hybridised to an array of sequences with known identities (probes). The RNA from the tissue of interest first needs to be isolated and purified (Fig. 4.1). It is important that this RNA be of a high quality with little degradation of the sample. RNA may need to be amplified to obtain amounts sufficient for experimentation and validation as the amount of RNA extracted from embryonic tissue may be limiting. The relatively unstable RNA transcript is copied to a more stable DNA molecule by using a special oligo (dT) primer (a single strand of poly T bases that contain a T7 promoter). This binds to the polyA sequence in the 3′ tail of mRNA molecules (Fig. 4.1). Using this primer and the mRNA template the complementary sequence of the mRNA molecule is made with reverse transcriptase enzyme, this is called complementary DNA (cDNA). The second strand of this cDNA is then made forming a stable double stranded DNA (dsDNA) molecule that retains the sequence information of the original mRNA molecule. Once this dsDNA has been purified it

is used as a template for the amplification process. *In vitro* transcription (IVT) is performed on the dsDNA template using the T7 RNA polymerase which binds to the T7 promoter incorporated into the dsDNA. During this IVT reaction thousands of RNA copies of each original mRNA transcript are made, these are known as amplified RNA (aRNA). During this process a modified uracil nucleotide, 5-(3-aminoallyl)-UTP (aaUTP) is incorporated into the aRNA. The aminoallyl modification of this molecule allows a dye label to be attached to the aRNA molecule in subsequent steps. This can be done with the commonly used cyanine 5 (Cy5) or cyanine 3 (Cy3) fluorescent dyes. When excited, these dyes fluoresce at different wavelengths. Cy5 is excited maximally at 649 nm and emits light maximally at 670 nm while Cy3 is excited maximally at 550 nm and emits light maximally at 570 nm. The ability to label aRNA samples from two different tissues with different dyes gives a distinct signal identity to each which makes them distinguishable from each other when combined (Shalon et al., 1996). This occurs in duel labelling experiments (when two samples are co-hybridised together), one RNA sample is labelled with Cy3 and the other with Cy5, the samples are then co-hybridised

Microarray hybridisations are performed on a solid-state surface to which probes are fixed. In this study 70mer oligonucleotide DNA microarrays, representing over 36 000 mouse cDNAs, were used and the discussion will therefore be limited to this type of platform. A DNA microarray platform essentially consists of a solid surface (conventionally a glass slide) onto which synthesised oligonucleotides (short sections of DNA with a sequence that corresponds to that of a specific gene) are fixed. These oligonucleotide sequences are the microarray probes. The oligonucleotides are 'spotted' onto the slide using a printing technique. Each spot that a pin prints on the array contains oligonucleotides with a sequence that corresponds to a specific gene. Many thousands of these spots can be printed onto one slide making it possible for the entire genome of an organism to be represented on one array.

In two-colour microarray experiments, two targets (each mRNA target sample being labelled with a different dye) are hybridised onto the same array, this is termed competitive hybridisation. During hybridisation each labelled mRNA sequence will anneal with the probe sequences that they are most complementary to.

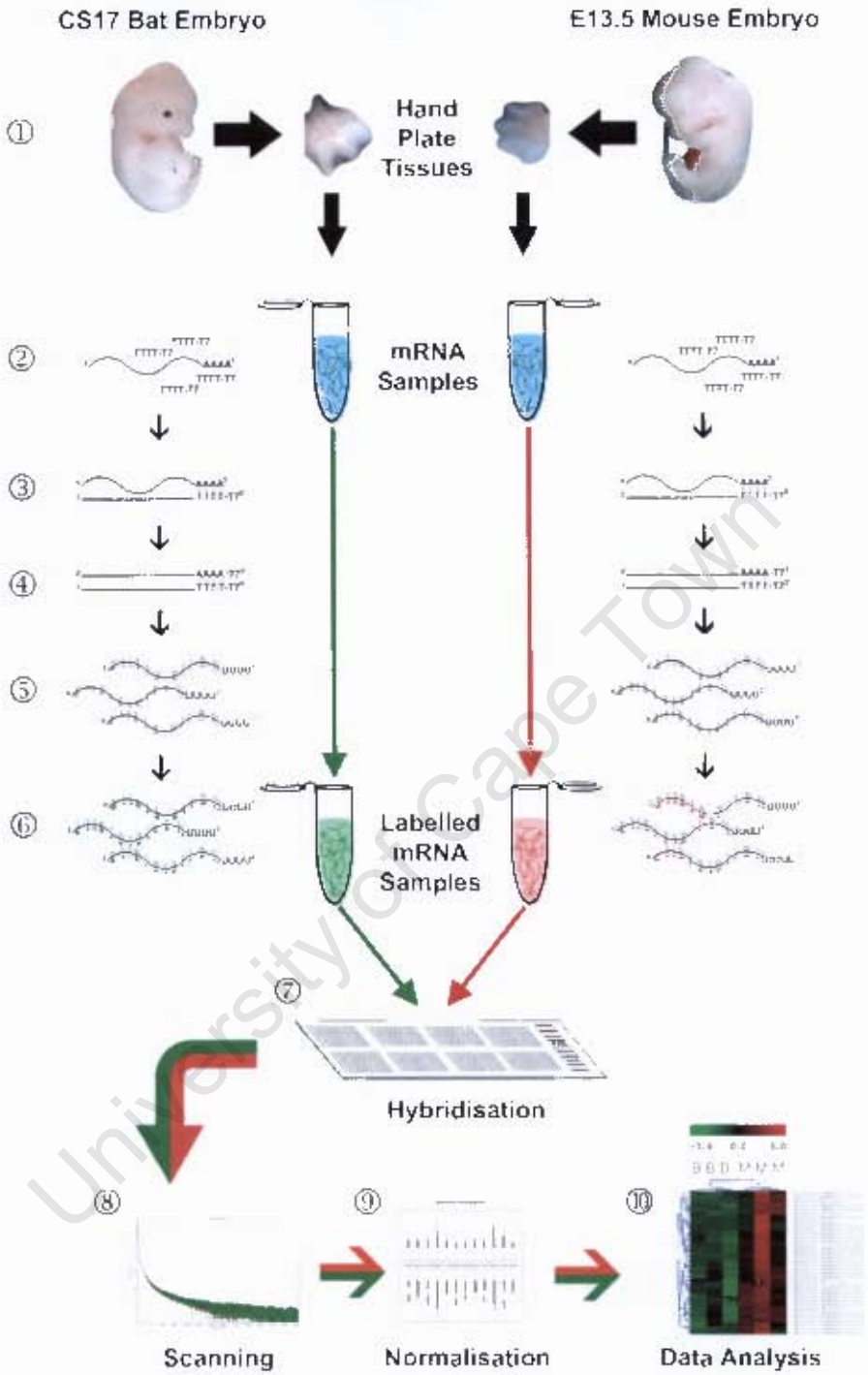


Fig. 4.1: Flow diagram showing a typical microarray experiment in which the hand plate of a CS17 bat embryo is hybridised against the hand plate of an E13.5 mouse embryo. In this experiment each tissue condition is processed individually. The RNA is first extracted and purified ①, an oligo dT primer ② is then used to make the cDNA ③. A second strand of the cDNA is then made forming a stable double-stranded molecule called dsDNA ④. This is used as the template for amplification. During amplification many copies of aRNA are made, these have incorporated dye binding sites ⑤. Fluorescent dye molecules attach to these sites during a labelling reaction, the reference and test samples are each labelled with a different dye ⑥. These two samples are then pooled together and hybridised on the same array ⑦. This array is then scanned, digitising the fluorescent signal that is emitted from each dye ⑧. The captured data is then normalised to remove technical variation ⑨ and then analysed using specifically designed software ⑩.

Conditions can be modified to promote specific hybridisation between labelled mRNA transcripts (other wise known as the targets) and the fixed oligo probes. After hybridisation, successive washing steps with increasing stringency are used to remove unbound target and target that is non-specifically bound to the probes. This reduces the false signals which arise from the non-specific binding that occurs during low stringency hybridisation.

Once slides have been hybridised and washed they are scanned (Fig. 4.1) to retrieve the fluorescent signal from each spot on the slide and capture it in a digital format. The signal from each dye is captured as an independent image, the signal from the Cy3 dye is captured as a green image (this can be referred to as the green channel) and that of the Cy5 is captured as a red image (this can be referred to as the red channel). The intensity from each photomultiplier tube (PMT) can be adjusted independently. In this way the signal from each channel can be adjusted so that it has an overall higher or lower intensity. As this adjustment of the PMT may introduce bias into the experiment it is important to standardise the settings used. This is usually done through making the overall signal from the red channel equivalent to that of the green channel and ensuring that there is a similar signal intensity distribution.

Data for each spot needs to be individually captured from these scans. However, though each spot representing each gene is visually distinct, the signal from each spot must be captured and made digitally distinct from the other features. This is done through a combination of computer algorithms that 'find' (recognise and isolate the spot from the background) and 'capture' (record the signal intensity from each channel as a value) the data from each spot. When spots are not distinct enough to be recognised by the algorithm they are flagged as 'not found'. The data is then double checked manually and features with erroneous signals (such as fluorescence caused by dust or wash effects) are flagged as 'bad'. This process of data capturing is termed 'spot-finding' and 'flagging'.

Each oligonucleotide spot on the spotted array has a specific gene identity that can be matched to the spot through the process of data capturing. In this way the signal that is captured from each spot will correspond to the mRNA abundance of a specific gene in the two samples being compared. The dual-labelling of the targets of interest allow the relative abundance of each of the target mRNA samples to be compared on a spot-by-spot basis. If there are only two tissues of interest then this can be done directly (both tissue samples are

hybridised on the same slide and compared). However, experimental design becomes important in experiments which are more complex, involving comparisons of more than two variables. The design refers to how the different targets are hybridised to the arrays to allow accurate comparisons between samples to be made, with the maximum amount of information gained from the experiment. Designs must also promote the minimum distance between comparisons to minimise the technical variability between experimental conditions. The design chosen for a particular experiment is dependant on: (i) the experimental question that is being asked; (ii) the type and number of samples that are going to be used to answer that question; (iii) the number of experimental and technical repeats that are needed; and (iv) the number of experiments that can be performed due to time constraint or budget limitations. The designs range from simple reference designs, where all the samples are co-hybridised with a common reference sample, to the more complicated loop designs, where samples are compared sequentially to one another (Churchill, 2002; Yang and Speed, 2002)

A microarray experiment is composed of a series of sequential steps. Each step (or process) may contribute in some way to the technical variation that can be found between samples, from biases in the RNA extraction step, to differences in dye incorporation and to differences in scanning and capture of data. The cumulative effect of these experimental variations can be quite large and needs to be minimised in a process known as data normalisation, prior to data analysis. Normalisation steps include background correction, within slide normalisation and between slide normalisation. Within array variations occur when the data on one array (probes present on one slide) are biased relative to one another. Between array variation occurs when the data between two or more arrays (a probe set across many different slides) are biased relative to one another.

After normalisation, the data from the microarray experiment can be analysed. Statistical tests are used to identify genes that are differentially expressed. It is assumed that genes that show differences in mRNA transcript abundance play an important role in the differences between the tissues being tested. In addition to identifying differentially expressed genes clustering algorithms can be used to group genes which show similar changes in mRNA transcript abundance. It is assumed that the expression of these genes will be co-ordinated. Functional annotation of these genes can be used to identify whether genes belonging to

particular biological, cellular or molecular functions are co-ordinately regulated in the tissues being compared

4.1.3 Cross-species microarrays and evo-devo

The completion of the sequencing of the genomes of model organisms has allowed the design of oligonucleotides which represent the full spectrum of all coding genes. These can be used in microarray experiments to characterise the transcriptomes of different tissues in model organisms under a range of different experimental conditions. This technology also has applications in the field of evolutionary biology (Chiu and Hamrick, 2002). Boutique arrays, which are based on available information from non-model organisms, can be designed and manufactured to allow species-specific hybridisations (SSH) to be made. However, if not commercially available, these arrays are expensive and time-consuming to manufacture and, may only contain a limited probe set. As a result, many species of interest in evo-devo research do not have a representative microarray platform for transcriptome analysis. An alternative approach is to do a Cross-Species Hybridisation (CSH). In this design, the target mRNA from the species of interest is hybridised to a microarray platform from a model organism (Bar-Or et al., 2007a). Cross species hybridisations are a non-standard use of microarray technology (Bar-Or et al., 2007a), however they have been validated as being a useful tool to compare evolutionarily similar species (Brodsky et al., 2005) and evolutionarily distant taxa (Adjaye et al., 2004; Grigoryev et al., 2005; Ji et al., 2004; Renn et al., 2004). This approach has proven useful in a variety of different studies involved in the evolutionary and developmental processes of a range of different organisms (Brodsky et al., 2005). However, it has been shown that bias can be introduced into the data due to species-specific differences of the mRNA targets (Cohen et al., 2007). For example, problems may arise when comparing relative mRNA transcript abundance in evolutionary distant species, such as that of the cow and the human on a human microarray platform (Fig. 4.2). These problems are due to increasing sequence dissimilarity with increasing phylogenetic distance. This occurs because microarray experiments are dependant on the hybridisation specificity between the labelled-mRNA sequence and the oligo sequence to create a signal on the slide.

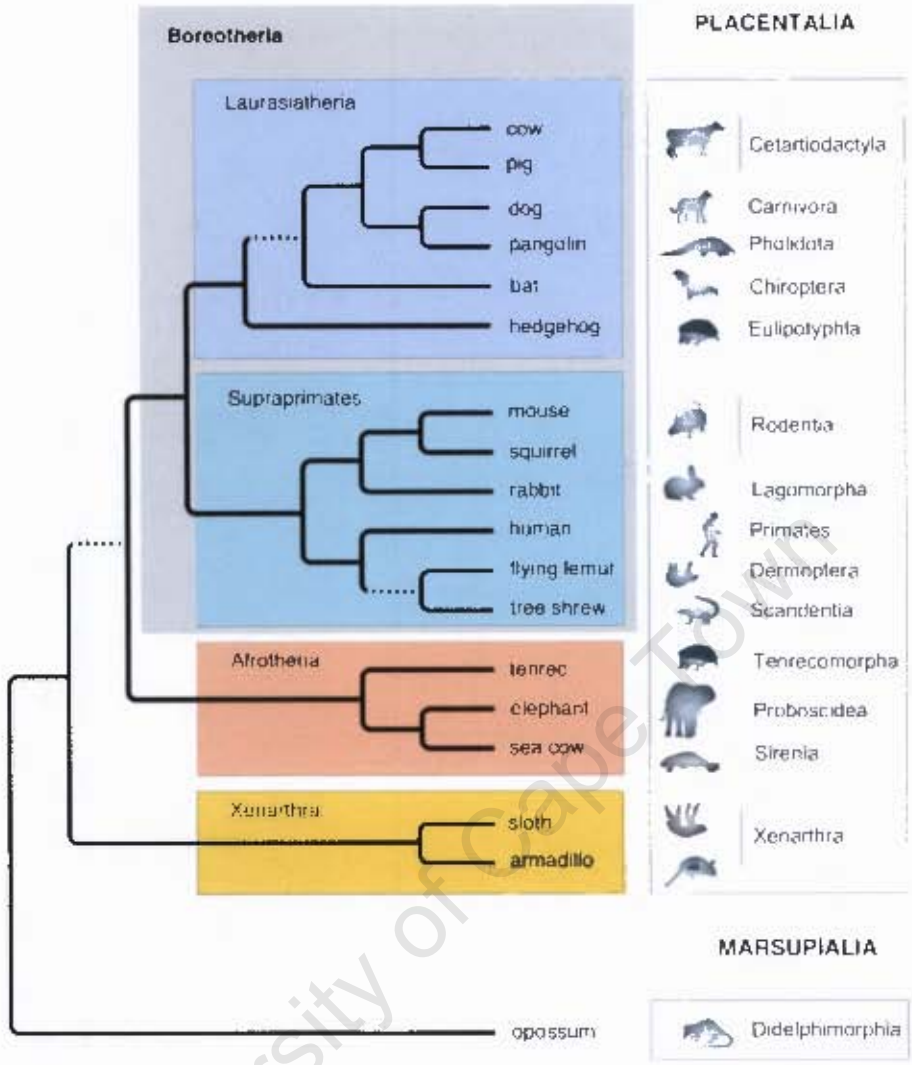


Fig. 4.2: Evolutionary tree showing the phylogenetic distance of placental mammalian species. The distance between the cow (order: Cetartiodactyla, clade: Laurasiatheria) and the human (order: Primates, clade: Supraprionates) is similar to that between the bat (order: Chiroptera) and the mouse (order: Rodentia). (Figure adapted from Kriegs et al. (2006) and Murphy et al. (2001)).

If the specificity of the labelled-mRNA is lower in one of the target species relative to another, the hybridisation efficiency of that species will be lower thus causing a lower signal on the array for that specific gene (Fig. 4.3). Thus the signal strength of certain mRNA transcripts may be affected by the similarity of their sequences in addition to their transcript abundance as compared to that of the oligo probe on the array (Bar-Or et al., 2007b).

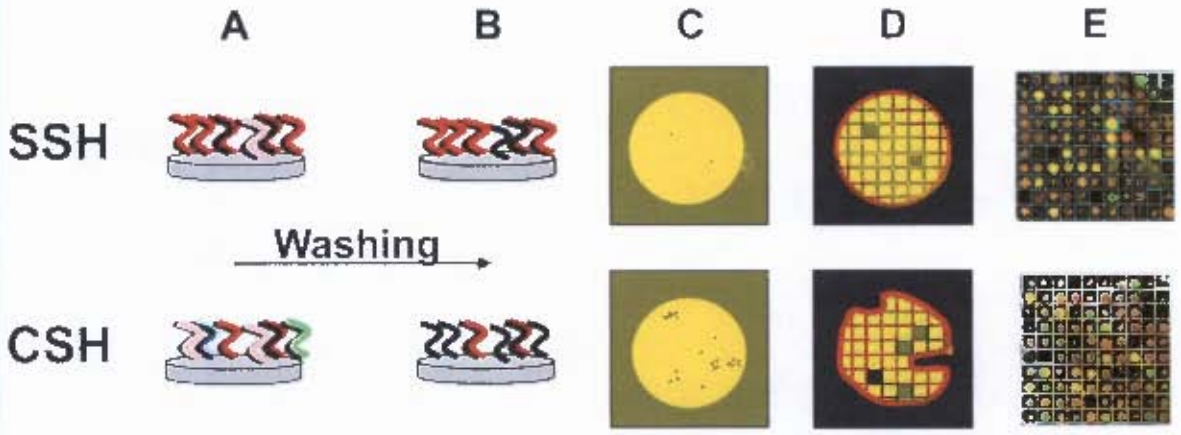


Fig. 4.3: Differences in hybridisation efficiencies can be seen between species-specific hybridisations (SSH) and cross-species hybridisations (CSH). After hybridisation (A) there are relatively more mismatches (pink strands) in the CSH. After the washing step, these mismatched pairs are removed (B) leaving only the perfect matches (red strands). This results in areas of weaker signal in some of the CSH spots (C) which, when captured result in low intensity pixels in the digital image creating a spot with an altered morphology (D). This also results in an overall decrease in the slide signal intensity (E) (adapted from Bar-Or et al (2007b)).

It is also expected that in CSH experiments there may be a higher level of mismatches, this occurs when mRNA transcripts of one gene bind to a probe that is intended to represent another gene (Bar-Or et al., 2007a). This again can be attributed to sequence divergence that occurs among the different species being hybridised. The lowered signal intensity for CSH (Fig. 4.3), has been attributed to this inefficiency of non-specific binding between the transcripts of the alternative species and the model organism probe set (Bar-Or et al., 2007b).

This problem is exacerbated by the fact that the oligo probes are preferentially designed to bind to the 3' untranslated (UTR) end of the labelled mRNA transcript, the sequence in this region tends to be less conserved among different organisms than coding regions. It has also been hypothesised that this non-specific binding in combination with subsequent high-stringency washing steps would contribute towards altering spot morphology in CSH (Fig. 4.3; (Bar-Or et al., 2007b)). As a result a novel method of filtering out this non-specific data based on several spot morphology criteria has been proposed (Bar-Or et al., 2007b). Cross species hybridisation experimentation and data analysis should be approached with caution and it is important to keep potential pitfalls and biases in mind; however, enough information exists about their validity for them to remain a relevant tool that enables the characterisation of previously inaccessible transcriptomes.

4.1.4 Microarray analysis of limb development in bats as compared to mice

There has been one previous study in which a bat CSH was performed. This was done on human cDNA glass arrays (19 000 genes) and sought to examine the gene expression within bat skeletal muscle both during and out of hibernation (Eddy and Storey, 2001). It was found that there was a high occurrence of false negatives that arose due to the low cross-reactivity of the certain mRNA targets to the probes. Thus only a subset of genes (15% of the 19 000 spot array) with an adequate signal was available for analysis. Of these genes five were found to be significantly up regulated in the skeletal muscle of hibernating bats as opposed to non-hibernating bats, (Eddy and Storey, 2001; Eddy and Storey, 2002). This study illustrates that, though a CSH might impose limitations due to the potential for false negatives, enough genes are available for analysis to enable biologically relevant information to be obtained, making this technique an important first step towards gene discovery.

In this study the transcriptome of embryonic bat limbs from CS17 and CS16 stages of development were analysed in conjunction with that of the mouse E13.5 hand plate. A mouse oligo array was used as there is currently no bat EST or oligo array available. The mouse handplate was included in the comparison as the current model of limb development is mainly based on the analysis of conventional model organisms such as the chick and the mouse. The mouse has also been used at the comparative model system in several previous bat limb developmental studies (Chen et al., 2005; Cretekos et al., 2007; Weatherbee et al., 2006) and has been posed as a good comparative model for the bat (Cretekos et al., 2001). Co-hybridisation of bat and mouse labelled-mRNA would give a relative measure of the gene expression that occurs between these two species. This would allow the bat limb development transcriptome to be directly compared to that of the mouse. The order Chiroptera (bat) and the order Rodentia (mouse) belong to two separate clades (Fig. 4.2) and are therefore not closely related. However, successful competitive CSH has been performed over a similar, or even greater phylogenetic distances (see Bar-Or et al. (2007a) for a review). The inclusion of 3 separate variables in the experiment resulted in a multi-dimensional dataset and allowed cross comparisons to be focused on differences between stages of development (CS17 and CS16), autopod type (bat hand and foot plate) or species-specific differences (E13.5 mouse hand plate and CS17 bat hand plate)

4.2 METHODS

4.2.1 Experimental design

The commonly used reference design was used in the microarray experiment, it is a robust design that allows simple and direct comparisons to be made between samples (Churchill, 2002; Kerr, 2003; Yang and Speed, 2002). In the experiment the mouse hand plate was used as the reference condition and the CS16 and CS17 bat hand and foot plates were the test conditions. Five experimental sessions were performed using OPERON Mouse OpArray slides. The bat RNA (the test) samples were labelled with Cy3 dye (green signal) which has a relatively strong and stable signal on arrays. The mouse RNA (the reference) sample was labelled with Cy5 dye (red signal) which is more prone to a loss of signal due to fluorescence quenching (Cox et al., 2004). This assignment of dyes was chosen as a relatively weak signal was already expected to occur from the bat labelled-mRNA sample due to cross-species hybridisation effects (Bar-Or et al., 2007a) therefore it would be preferential for any signal dampening that occurs due to dye effects to take place in the channel of the mouse labelled-mRNA (species-specific sample).

4.2.2 Limb samples

Bat embryos used for RNA extraction were obtained and stored in RNA*later*® solution using methods described in Section 2.2.2. Four CS16 embryos and four CS17 embryos were used in the subsequent experiments. These stages were confirmed through examination of the embryos both before and after they had been stored in RNA*later*®.

The mice chosen for this study were an out bred strain (ICR), which has inherent genetic variability. They could therefore be used in a comparison with bats which are non-captive, genetically variable animals. Mouse embryos were obtained from the Animal Unit at the University of Cape Town. Ethics approval to mate and sacrifice mice was obtained from Animal Research Ethics Committee at the Faculty of Health Science, University of Cape Town (application number: 006/040). Timed matings were performed and on day 13.5 the pregnant female mice were euthanized using a 5 min exposure to halothane followed by cervical dislocation. The embryos were then dissected out. Those used for RNA extraction were stored in approximately 10X their volume of RNA*later*®. These were put at 4˚C for 24 hrs and then stored at -20˚C; samples were kept at -80˚C for long-term storage. A total of

twenty-four E13.5 mouse embryos, combined from three litters at the equivalent stage of development were obtained for this experiment. These were pooled to create a large reference pool that would be sufficient for all the microarray experiments and downstream analyses. One representative embryo from each litter was fixed in 4% PFA at 4˚C for 24 hrs before dehydrating them in methanol washes, and storing them in 100% methanol at -20˚C. This allowed the mouse developmental stages obtained from each pregnancy to be confirmed using the limb staging system of Wanek et al. (1989) and the Mouse Atlas of Development (Kaufman, 1992).

The autopods (the hand and foot-plate area) of the bat and mouse embryos were dissected out, using an SZ51 stereomicroscope (Olympus UK Ltd., Waltford, UK), whilst the embryo remained immersed in RNA*later*®. The autopod area was distinguished as being the area of dorso-ventrally flattened tissue on the distal portion of the limb. The autopods were cut off along the point of constriction of the presumptive wrist or ankle excluding the tissue associated with the stylopod and the zeugopod as well as the tissue associated with the proto- and plagiopatagium (Fig. 4.4).

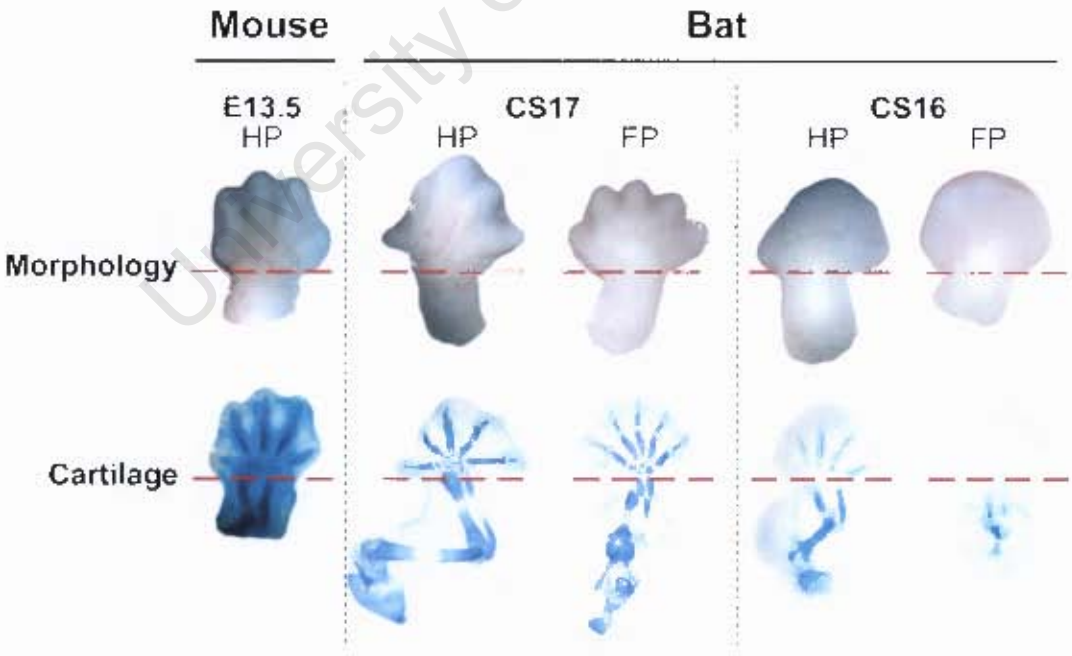


Fig. 4.4: Limb samples showing morphology, cartilage and approximate region at which the hand and foot plate were cut off. Images are not to scale (cartilage stained mouse image from Reno et al. (2008), cartilage stained bat images courtesy of R. Behringer)

The autopods of the bat samples were stored separately (with the left and right hand side being combined) in fresh solutions of RNA*later*®. All the forelimb autopods of the mouse were combined and stored together in a fresh solution of RNA*later*®. These samples were stored at -80°C.

4.2.3 RNA extraction and quality check

RNA from the autopod tissues was extracted using the RNeasy® Lipid Tissue Mini Kit (QIAGEN, Valancia, CA, USA). All implements used to handle the samples were soaked overnight in 0.4M sodium hydroxide (NaOH) solution and then rinsed with DEPC treated water. The autopods of the mouse embryos were pooled into two groups (one with eleven pairs of hand-plates and one with ten pairs). The RNA was extracted from these two groups on separate columns in a paired extraction and later pooled together. This was done to ensure that the capacity of the column (100 mg of tissue) was not exceeded. The left and right-hand tissues of each bat limb condition (hand plate or foot plate) were pooled together and processed as a pair. The tissue samples were removed from the RNA*later*® solution and placed in 0.5 µl QIAzol Lysis Reagent. The tissues were homogenised using a plastic microfuge tube pestle attached to a Bosch PSB 400-2 drill (Robert Bosch (Pty) Ltd., Midrand, South Africa). The material was homogenised for a minimum of 5 min after-which an additional 0.5 µl of QIAzol Lysis Reagent was added. The sample was homogenised for a further 5 min and then left to stand at room temperature for a minimum of 5 min. After this 200 µl of chloroform was added to the homogenate and the tube was shaken vigorously for 15 s. The homogenate was left to stand at room temperature for 3 min and was then centrifuged at 12,000 x *g* for 15 min at 4°C. This separated the homogenate into three phases: (i) an upper, colorless, aqueous phase containing RNA; (ii) a white interphase; (iii) and a lower, red, organic phase. The upper aqueous phase was transferred to a new collection tube and one volume of 70% ethanol was added. This was vortexed and half of the sample was loaded onto an RNeasy Mini Spin Column in a 2 ml collection tube. This was centrifuged at 8000 x *g* for 15 s at room temperature and the flow through discarded. This step was repeated with the rest of the sample. Next, 700 µl Buffer RW1 was added to the column and centrifuged at 8000 x *g* for 15 s, the flow thorough was once again discarded. The column was transferred to a new collection tube and 500 µl Buffer RPE was added. This was centrifuged at 8000 x *g* for 15 s and the flow through discarded. This step was repeated and the column centrifuged for 2 min. The RNeasy Spin Column was placed into a new collection tube and centrifuged at full speed for 1 min to dry the silica-gel membrane. The column was transferred to a new collection

tube and 50 µl RNase-free water was loaded directly onto the RNeasy silica-gel membrane. This was centrifuged at 8000 x *g* for 1 min to elute the RNA bound to the membrane. This step was repeated with a second volume of RNase-free water. .

The quality of the RNA from the mouse hand plate pooled reference and ten randomly chosen bat limb samples was checked by the Centre for Proteomic and Genomic Research (CPGR, Cape Town, South Africa) using an Agilent 2100 Bioanalyser (Agilent Technologies, Stockport, UK). The RNA integrity of the other six bat limb samples was checked by loading 1 µg of RNA onto a 2% denaturing agarose gel run at 80 V for 1 hr. The concentration and condition of the sample was determined using the Nanodrop® ND-1000 Uv-Vis Spectrophotometer (Nanodrop Technologies, Inc., Wilmington, DE, USA) prior to the experiment. These measurements were repeated three times and the average calculated. The full absorbance profile of each sample was visualised between 220 nm and 320 nm and the 260/280 (A_{260}/A_{280}) absorbance ratio was calculated. Samples were stored at -80°C.

4.2.4 Amplification and labelling

The RNA samples were amplified and labelled using the Amino Allyl MessageAmp™ II Cy3 aRNA Amplification Kit (Ambion, Austin, TX, USA) as per manufacturer's instructions. A fixed amount of 0.5 µg of input RNA was used in all experiments.

First strand cDNA synthesis was performed using T7 oligo (dT) primers, RNase inhibitor and Arrayscript™ reverse transcriptase as per manufacturer's instructions. Each RNA sample was combined with 1 µl of the T7 oligo (dT) primers and incubated in a Genius Thermal Cycler, model FGN02CD (Techne, Cambridge, UK), for 10 min, allowing the primers to anneal to the polyA tail of the RNA molecules. A reverse transcriptase master mix, composed of 2 µl 10X first strand buffer, 4 µl dNTP mix, 1 µl RNase inhibitor and 1 µl Arrayscript™ reverse transcriptase, was added and each sample incubated in a Series 2000 digitally controlled oven, model 275 (Scientific, Industria, South Africa) at 42°C for 2 hrs.

A second strand master mix was prepared on ice; it was composed of 63 µl Nuclease-free Water, 10 µl 10X Second Strand Buffer, 4 µl dNTP Mix, 2 µl DNA Polymerase and 1 µl RNase H. A Maximum Plus™ vortex (Thermoline Scientific, Smithfield, NSW, Australia) was used to mix samples when required. This master mix was added to each 20 µl sample,

making them up to 100 µl. These samples were incubated in the thermal cycler at 16 °C for 2 hrs.

The synthesised cDNA was then purified. Each sample was mixed with 250 µl cDNA binding buffer and loaded onto a cDNA filter cartridge. This was centrifuged at 10 000 x *g* for 1 min using a Force 1418 bench-top centrifuge (Labnet International, Inc., Windsor, Berkshire, UK) and the flow through discarded. This was repeated with 500 µl wash buffer loaded onto each column. The filter cartridge was transferred to an elution tube and 9 µl of nuclease-free water, that had been heated to between 50 and 55 °C using a Stuart® SBH130 block heater (Barloworld Scientific, Stone, Staffordshire, UK), was loaded. Samples were left to stand at room temperature for 2 min and centrifuged at 10 000 x *g* for 1 min 30 s. This step was repeated with a second 9 µl of preheated nuclease-free water and the elution, containing the purified cDNA, was set aside.

In Vitro Transcription (IVT) was then performed on the purified cDNA to synthesise aRNA. In this reaction an IVT master mix, composed of 3 µl aaUTP (50 mM), 12 µl ATP, CTP, GTP Mix (25 mM), 3 µl UTP solution (50 mM), 4 µl T7 10X reaction buffer and4 µl T7 enzyme mix, was added to each purified cDNA sample. This gave a 1:1 ratio of modified aminoallyl UTP (aaUTP) nucleotides to unmodified UTP nucleotides. These reactions were incubated in a Series 2000 oven, model 275 (Scientific, Industria, RSA) at 37°C for 16 hrs and stopped by the addition of 60 µl Nuclease-free Water to each aRNA sample.

Each sample of synthesised aRNA was mixed with 350 µl aRNA binding buffer and 250 µl 100% ethanol and immediately loaded onto an aRNA filter cartridge. The samples were centrifuged at 10 000 x *g* for 1 min and the flow through discarded. This was repeated with 650 µl wash buffer loaded onto each column and the column was spun dry for an additional minute. The filter cartridge was transferred to an elution tube and 100 µl of nuclease-free water, which had been heated to between 50 and 60 °C, was loaded. Samples were left to stand at room temperature for 2 min and centrifuged at 10 000 x *g* for 1 min 30 s. The elution, containing the purified aRNA, was quantified using the Nanodrop® ND-1000. The quality of the purified aRNA was checked by loading 1 µg of aRNA onto a 2% denaturing agarose gel run at 80 V for 1 hr.

The samples were labelled with cyanine fluorescent dyes. Photobleaching of the samples was prevented by handling these dyes under minimal light conditions and protecting the samples from light at all times. The dyes were re-suspended in 11 µl of Dimethyl sulfoxide (DMSO) and left to stand for an hour. Each aRNA sample (20 µg) was dehydrated using the Savant Speedvac® plus (SC210A) attached to a refrigerated vapour trap (RVT400) powered by a Valupump (VLP120) (Thermo Scientific, Milford, MA, USA). The speedy-vac was set on low (i.e. the samples were not heated) and all samples took less than thirty minutes to become fully dehydrated. The remainder of the purified aRNA sample was stored at -80 °C. The dehydrated samples were re-suspended in 9 µl coupling buffer and the appropriate dye was added. The reference and test samples were labelled with Cy5 and Cy3 dyes (Amersham Biosciences, Piscataway, NJ, USA) respectively. This mixture was left to stand for 30 min. The reaction was quenched by adding 4.5 µl 4M hydroxylamine and leaving this mixture to stand for 15 min. After this 5.5 µl nuclease-free water was added to each sample to bring the volume to 30 µl. Each sample of labelled aRNA was mixed with 105 µl aRNA binding buffer and 75 µl 100% ethanol and immediately loaded onto an aRNA filter cartridge. The samples were centrifuged at 10 000 x *g* for 1 min and the flow through discarded. This was repeated with 500 µl wash buffer loaded onto each column and the column was spun dry for an additional minute. The filter cartridge was transferred to an elution tube and 10 µl of nuclease-free water, which had been heated to between 50 and 60 °C, was loaded. Samples were left to stand at room temperature for 2 min and centrifuged at 10 000 x *g* for 1 min 30 s. This step was repeated with a second volume of heated nuclease-free water. The elution, containing the purified labeled aRNA, was quantified using the Nanodrop® ND-1000. The labelled aRNA sample was diluted 1:3 with nuclease-free water to increase the loading volume to one which could be accurately pipetted. The dye incorporation and aRNA quantity of the diluted sample was also measured on the Nanodrop® ND-1000.

4.2.5 Slide preparation

The *Mus musculus* (mouse) OpArray™ was the platform used in this experiment. This platform consists of activated epoxide coated slides printed with version 4.0 of the *Mus musculus* (mouse) Array Ready Oligo Set (AROS), which contains over 36 000 cDNA probes (OPERON Biotechnologies, Inc., Huntsville, AL, USA). These Operon OpArray™ slides were printed by Operon and were treated according to the OpArray™ Protocol (Operon, 2007a). During pre-hybridisation the slides were placed in a sterile slide holder and submerged in OpArray Pre-Hyb Solution that had been preheated to 42 °C. This was

incubated in the oven for 1 hr at 42 °C. The slides were placed in a slide rack, washed in a premixed OPERON wash solution for 5 min and rinsed in MilliQ water three times for 30 s each. These washes were done at room temperature for 30 s each on an Orbital shaker, model os-20 (Boeco, Boekel + Co. (GmbH +Co) Hamburg, Germany), set on low. Transfer procedures were done quickly and the slides were not allowed to dry in between washes. After washing the slides were immediately transferred to a dry slide rack and dried by centrifugation in a Sigma Laboratory Centrifuge, model 4-15C (SIGMA Laborzentrifugen GmbH, Osterode am Harz, Germany) at 200 x *g* for 5 min. Slides were used immediately after pre-hybridisation and washing. Prior to hybridisation the LifterSlips™ (Erie Scientific Company, Portsmouth, NH, USA) were immersed in acetone while being agitated on an Orbital shaker, model os-20 (Boeco, Boekel + Co. (GmbH + Co.) Hamburg, Germany) for an hour. They were washed in a 0.2% SDS solution for 10 min and then rinsed in water, for 10 min, twice. The clean LifterSlips™ were dried in the oven at 50 °C for a minimum of 30 min.

4.2.6 Hybridisation and washing

Prior to hybridisation target Cy3-labelled aRNA sample was combined with the reference Cy5-labelled aRNAsample. A total of 250 ng of each labelled aRNA sample was added to 45 µL of OpArray Hyb Buffer. This ensured that in all the experiments the total concentration of Cy5-labelled target or Cy3-labelled target did not exceed the recommended concentration of 0.8 pmol/µl. The target was denatured at 65 °C for 5 min, centrifuged and loaded immediately or placed on ice. The dried, pre-hybridised slides were placed on a solid heating block set at 50°C. The LifterSlip™ was placed, white edge down, over the array creating a hybridisation chamber. The target solution was slowly pipetted along the bottom edge of the LifterSlip™ allowing the solution to be evenly wicked over the array by capillary action. The slide was placed in hydrated hybridisation cassettes, which were sealed and incubated in the dark in a water bath at 42 °C. Slides were incubated for 19 hrs.

Slides were washed in three successive wash solutions of increasing stringency. The three OPERON slide set washes were premixed, wash 1 consisted of 2X SSC and 0.1% SDS that was preheated to 42 °C, wash 2 was 1X SSC and wash 3 was 0.2X SSC. The slides were transferred to a slide rack and gently agitated using an Orbital shaker in the first wash solution for 10 min. They were transferred to a clean slide rack and agitated in the second and third wash solution for 5 min each. After washing slides were immediately transferred to a dry slide rack and dried by centrifugation at 200 x *g* for 5 min. They were kept in the sealed

centrifuge until scanning which was done within 6 hrs. After scanning the slides were placed in sealed, light free slide holders.

4.2.7 Scanning and data capture

The slides were scanned using a Genepix® 4000A scanner (Axon Instruments Inc., Molecular Devices, Sunnyvale, CA, USA). A preview scan was performed, this is a low resolution (40 µm/pixel) scan used to optimise the scanner settings without photo bleaching the dye and reducing its signal. The gain of the Photomultiplier Tubes (PMT) for each wavelength of light emitted was adjusted to balance the signal from each channel. The histogram and the count ratio (CR) of the preview scan were examined. The signals from each channel were standardised by adjusting the PMTs until the count ratio was a value between 0.95 and 1.05 and the histogram of both channels had similar (overlapping) intensity profiles.

The data was captured and flagged using GenePix™ Pro ver. 6.0 microarray analysis software (Amersham Biosciences, Piscataway, NJ, USA). The data associated with each specific spot image was captured and linked to its Operon oligo identity number using the *Mus musculus* (mouse) OpArray Gal file which can be downloaded from the OpArray™ Resource File Downloads page (Operon, 2007b). In this procedure the spot image (known as a feature) that has been captured digitally through scanning is discriminated from the background. This is done by aligning a feature indicator over the feature, allowing the several measurements associated with that specific feature to be taken. Spots were captured automatically as irregular features rather than circular ones. This allows a better discrimination between their morphologies (i.e. spot circularity), which can be used as a flagging index for CSH datasets (Bar-Or et al., 2007a). Features were allowed to be resized from a maximum of 300% to a minimum of 33% of their original size (±100 µm).

The background signal was calculated using the method of morphological closing followed by opening. This is a mathematical morphological technique that estimates the background intensity of each channel by copying the original image and then filtering it. In the first step (closing) small dark areas in the image are filled in, in the second step (opening) a local minimum filter is applied to the whole image (Axon Instruments Inc., 2004). In essence a slide image is created in which all the spots and any particle fluorescence has been removed, this is done by applying a local background minimum reading that leaves out bright areas and anomalous small dark regions (see Bengtsson (2003) for more details). This method produces

a background that is slightly lower in areas of low background intensity and significantly lower in areas with high background intensity (Axon Instruments Inc., 2004). Features that failed to be discriminated from the background, that were smaller than six pixels or that fell outside of the feature resizing threshold were automatically flagged as "Not Found" and given a flag value of -50. Found features were given a flagging value of 0. Features that had been captured automatically were checked manually and any automatic capture errors were corrected. During this time anomalous features were manually flagged as bad. These included spots that were discoloured (over green) due to obvious wash effects; spots that were marked by the fluorescence of dust particles, hair or a scratch; and spots that had merged due to printing errors. Dye marker spots that occurred at the top, left-hand corner of each block were also marked bad manually. Features were then automatically flagged as bad using the 'Flag Features' function. Spots were flagged as bad if they were: over saturated (if they were greater than or equal to 10% saturated in either channel); misshapen (with a circularity of less than 35); if they were very small (with a diameter of less than 80 or a number of pixels less than 40); or if the spot signal was inconsistent (if the coefficient of variation of the pixels in either channel exceeded 99). These parameters were determined by visual inspection of the scatter plots of these variables for each array in Genepix™ Pro. The latter three parameters exclude spots based on their morphology and the consistency of the signal across the spot, which is a recommended method to be used when performing competitive CSH (Bar-Or et al., 2007a). The raw data was then exported from Genepix™ Pro as Genepix results (.gpr) files.

4.2.8 Background correction and normalisation

Background correction, normalisation and subsequent preprocessing steps for the OPERON arrays were done using the R programming language ver. 2.6.0. (R Development Core Team, 2008) using R Bioconductor ver. 2.1. (Gentleman et al., 2004) packages: biobase ver. 1.16.1.(Gentleman et al., 2006); limma ver. 2.12.0. (Ritchie et al., 2007; Smyth et al., 2005; Smyth and Speed, 2003); marray ver. 1.15.1. (Yang, 2008); convert ver. 1.11.0.(Smyth et al., 2004); stats ver. 2.6.0 (R Development Core Team, 2008); nlme ver. 3.1-85 (Pinheiro et al., 2007); agce ver. 1.2 (Gottardo, 2008); impute ver. 1.10.0 (Hastie et al., 2007); annotationTools ver. 1.8.0 (Kuhn, 2007), and a modified version of yasma ver. 0.20 (Wernisch, 2003), that excluded background subtraction steps in its functions. The median foreground signal intensity of each spot for each of the channels, Cy5 (red) and Cy3 (green) as well as the median background intensity of each spot for each channel was imported into R.

The annotation information weights matrix, with all spots flagged as bad (< -75) being weighted as 1 (Appendix A – R Script: Step 1).

The 48 tiling spots (dye markers) spotted on each array have no inherent biological information, and were removed (replaced with NA) prior to normalisation to prevent their high intensities and biased ratios (Section 4.3.5) from affecting the normalisation procedure. Removal of these spots left 37 392 spots on the array; this was then considered as the total number of spots on the array (Appendix A – R Script: Step 2). The distribution of the estimated background (made up of 37 392 spot value parameters) was compared to three spot groups that can be used as indicators for low to no signal on the OpArray slides. These are: null spots (154 spot areas per array that are not printed), buffer spots (1 006 spots per array that are only printed with buffer solution) and negative controls (224 spots per array that are printed with a probe that should not bind to mRNA transcripts that occur in the mouse transcriptome). The A values (Eq. 4.1) were used as an indicator of the spot signal intensity value, as its value represents the combined signal from each channel. The A value distributions of these groups was visualised using box plots and summary statistics were calculated.

$$A = \log_2\left(\frac{Cy3\ (green)\ signal + Cy5\ (red)\ signal}{2}\right) \qquad \text{(Eq. 4.1)}$$

The *normexp* background correction method was used with an offset of 32 to correct for any spatial artefacts that were present on the array. Data were adjusted within each array using a robust-spline normalisation function that was modified to accept and increased number of iterations (Boutros, 2004). Data were adjusted between arrays using R-quantile normalisation which does not exclude weighted values (Yang and Thorne, 2003). Data was visualised using density plots, boxplots of M (Eq. 4.2) and A values and MA plots (Appendix A – R Script: Step 3).

$$M = \log_2\left(\frac{Cy5\ (red)\ signal}{Cy3\ (green)\ signal}\right) \qquad \text{(Eq. 4.2)}$$

Density plots plot the frequency of the $\log_2$ intensity values of both the red and the green channel of each array separately; they allow the distribution of the different channels and arrays to be visualised. Boxplots summarise the density distribution of each array. They give

information on the median (centre bar in box); the first and third quartile (lower and upper hinge of box); the calculated minimum, first quartile + 1.5 x the inter-quartile range; the maximum value, third quartile + 1.5 x the inter-quartile range (lower and upper whisker); and the outliers (plotted as points) of each array. MA plots show the relationship between the M and A values, by plotting the log ratio data (M values) against the average intensity data (A values).

4.2.9 Filtering and preprocessing

An estimated threshold for low signal was calculated using the distribution of the normalised values for the null, buffer and negative control spots (Appendix A – R Script Step 4). This was done by calculating the maximum value for each of these groups (third quartile + 1.5 x the inter-quartile range) and taking the average of these. The A value threshold was 7.1. This is equivalent to an average unlogged intensity of 137. Genes that were weighted bad more than four times across all arrays were also removed. This filtering step was performed to ensure that poor quality data was removed (Appendix A – R Script: Step 5).

Values that were weighted were removed (replaced with missing values – NA). Probe replicates were merged by averaging the M values and A values by array (missing values were excluded in this process). Missing values were imputed using KNN imputation; this needed to be done to fulfil the requirements for a complete dataset in downstream data processes. KNN imputation takes advantage of the correlation structure inherent in microarray experiments to provide a robust and accurate estimator of missing values (Troyanskaya et al., 2001). This function selects a number of genes (specified by the parameter "k") with the most similar pattern to that of the gene with the missing value, using a Euclidian metric to calculate distance. The missing value is then replaced with the average of all the values of the selected genes for that column. The complete data set, post filtering, was used to determine the most appropriate value for k. Randomly generated missing values were assigned to 6% of this data set to create a test dataset. Imputation at increasing values of k was performed on this dataset. The normalised root mean square (NRMS) error (θ) was calculated using Equation (Eq. 4.3.

$$\theta = \frac{(M - M_{imp})_{RMS}}{M_{RMS}} \qquad \text{(Eq. 4.3)}$$

where M is the original dataset (test dataset without missing values) and M_{imp} is the dataset in which the randomly generated missing values have been imputed (Sehgal et al., 2005; Troyanskaya et al., 2001). The value of k that minimised the NRMS error was considered to be the most appropriate for imputation. The NRMS was found to be robustly minimised when k ranged from 15 to 20, k was therefore set at 18 when imputing missing values for this dataset (Appendix A – R Script Step 6).

To prevent batch effects from confounding data analysis they were corrected for using the R script for COMbating BATch Effects (COMBAT) (Johnson et al., 2007). This method relies on an empirical Bayes framework to estimate batch effect parameters from the gene-wise standardised data. The original data is then adjusted using these parameters. This method has been used to successfully correct for batch effects in a variety of recent studies (Acharya et al., 2008; Anders et al., 2008; Martínez-Llordella et al., 2008). To assess the effectiveness of batch correction on the different array condition groups the standard deviation within these groups was calculated both before and after batch correction and then plotted. To ensure that batch correction did not result in an overall reduction in variance between the condition groups the standard deviation of the mean of all these groups was calculated both before and after batch correction and then plotted (Appendix A – R Script: Step 7).

An additional filtering step was performed to reduce the number of genes tested for differential expression, this step removed genes which showed a consistent signal across all arrays (flat patterns). This was done by calculating the standard deviation (SD) of the signal for each gene across all arrays and excluding those genes whose SD fell below a specific threshold. This threshold was set to be the average SD of all the genes Appendix A – R Script: Step 8).

The relationship between the different arrays was visualised using Principal Components Analysis (PCA) after each step of the normalisation and pre-processing pipeline (Appendix A – R Script: Step 9). PCA was performed using single value decomposition of the centred and scaled data matrix. Screeplots, PCA plots of the first two components and hierarchal clustering trees (using Pearson correlation matrix to calculate distance and complete linkage when clustering) were used to visualise the relationship between the arrays.

4.2.10 Data analysis

The filtered and pre-processed data was analysed in R using the package DEDS ver. 1.12.1. This package looks for differential expression via distance synthesis (DEDS) of related measures. It integrates information across multiple ranking statistics and creates a robust estimate to rank differentially expressed genes (Xiao and Yang, 2007; Yang et al., 2005). The statistics used within the DEDS function were the fold change (FC), moderated *t* (mod*t*) statistic and significance analysis of microarrays (SAM). FC is commonly used to compare the mRNA expression level of a gene between two conditions (Cui and Churchill, 2003). In these analyses the DEDS package calculates the $\log_2$(FC). For a direct comparison this is the average M value of the biological repeats, however for an indirect comparison this is the difference between the average M values of the two conditions. The FC value was calculated by taking the antilog of this value. The mod*t* statistic and SAM are both modifications of the standard *t*-statistic. This statistic is calculated as the ratio between the difference of the means of the two comparison groups and their standard error (SE). Mod*t* and SAM differ from the standard *t* statistic by the addition of a constant to the estimate of the SD used in the calculation of the SE, this serves to regulate the statistic by increasing its denominator. In the mod*t* method this constant is chosen as that which minimises the coefficient of variation of the *t* statistic, while in the SAM method it is assigned in an *ad hoc* manner (Yang et al., 2005). This adjusts for the high amount of variation that is inherently incorporated into the estimation of the standard deviation in microarray experiments. This variation occurs due to the high number of genes and low number of replicates in these types of experiments. This modification makes the *t* statistic a more robust estimate of rank for experiments with a low number of replicates (Smyth, 2004). The use of both the mod*t* statistic and SAM may have introduced some redundancy in ranking the genes however this has been shown not to negatively affect DEDS analysis (Yang et al., 2005). Once genes were ranked according to their level of differential expression, the threshold of significance above which that gene is determined to be differentially expressed was determined. This significance threshold needed to factor in the problems associated with multiple testing that are inherent in all microarray analyses. This is because in these types of analyses thousands genes are being tested for significance simultaneously, which, increases the probability for accumulating Type I errors (making a false positive – declaring a gene as significantly expressed when it is not). This was taken into consideration by calculating the Benjamini and Hochberg False Discovery Rate (BH-FDR) for each gene (Benjamini and Hochberg, 1995). The BH-FDR of a gene,

given by the q value, is an indicator of the proportion of genes that represent false positives in the list of significant genes, if that gene is used as the threshold for significance (Cui and Churchill, 2003). This is a less stringent and more flexible method than most other multiple correction methods as, rather than controlling for any false positives, it allows you to control the proportion of false positives that you are willing to accept

Data analysis followed two separate pathways (Appendix A – R Script: Step 10): the first examined the differences between each of the four bat limb conditions and the mouse limb condition (bat-mouse comparison) while the second looked for differences between the bat limb conditions (bat-bat comparison). For the bat-mouse comparison four direct one-sample analyses for the arrays belonging to each limb condition group were done. In these analyses each bat limb condition group was compared to their corresponding mouse reference group. For the bat-bat comparison two indirect two-sample analyses were done. In these analyses the bat hand plate was compared to the bat foot plate of the corresponding stage. For all analyses genes were ranked according to the cumulative q value calculated by DEDS. Genes were determined to be differentially expressed if no more than one percent of the list of significant genes were likely to be false positives ($q < 0.01$).

In the bat mouse analysis these lists of significant genes were found to be over populated and were filtered to exclude those that were less than two fold over- or underexpressed. Gene list redundancy across the tissue types was then used to exclude species-specific gene expression (those genes that were over- or underexpressed in both the bat hand plate and the foot plate of the specific stages as compared to the mouse hand plate). In the bat-bat analysis these lists of genes were found to be sparse due to the limitations imposed by multiple testing (Section 4.3.9). A biological filtering step was performed on the top five percent of genes for each list. In this step genes that were over- or underexpressed more than 1.5 fold in the hand or in the foot plate of both bat stages were isolated.

Functional information was extracted from these gene lists using the online program Fatiscan (Al-Shahrour et al., 2006). Gene names were ordered by the moderated *t* statistic and tested to see whether gene annotations were over- or underrepresented within the comparison. This was done in Fatiscan by partitioning the data into blocks of genes (set to 60) and then testing

whether the distributions of functional annotations (GO molecular and biological processes) were asymmetrically distributed among these partitions (Al-Shahrour et al., 2006).

Data for genes of interest were extracted as single channel signals and plotted using Microsoft Excel 2007. The gene expression (signal) was taken as the average signal across all four biological replicates and the standard errors (SE) were calculated. Three comparisons were tested to see if they were significantly different using the Mann-Whitney U test in Statistica ver. 8.0 (StatSoft Inc., 2007). The mouse HP was compared to the CS17 HP; the CS17 HP was compared to the CS17 FP; and the CS16 HP was compared to the CS16 FP.

4.2.11 Bioinformatics

Gene annotation of unidentified probes was done using the NCBI Blast tool (Zhang et al., 2000) on the nucleotide collection database (nr/nt), with searches optimised for highly similar sequences (megablast). Any sequences matches with an E value of less than 0.05 were considered to be not significantly similar. Sequence homologies between the OpArray mouse probe and *Myotis lucifugus* cDNA were calculated for probes which were identified as representing differentially expressed mouse transcripts. The corresponding mouse and bat cDNA sequences were obtained from the Ensembl genome browser release 50. Sequences and annotation information for the *Meis2* gene was obtained from the Ensembl genome browser release 52 (Birney et al., 2004). Sequences were aligned together with the probe using the Bioedit Sequence Alignment Editor ver. 7.0.9.0. Gene–probe similarity was calculated as the percentage of matched nucleotides over the binding site of the OPERON oligo probe (Hall, 1999). If a probe had less than eighty percent similarity to the bat sequence, it was blasted using the NCBI Blast tool on the nucleotide collection database to see if the probe would hybridise to any other bat sequence. The blast was therefore limited to search within the taxa Chiroptera (taxid: 9397) with searches optimised for somewhat similar sequences (blastn). Any sequences matches with an E value greater than 0.05 were considered to be not significant. Sequences with corresponding annotation information for the *Meis2* gene was obtained from the Ensembl genome browser release 52 (Birney et al., 2004).

4.3 RESULTS

4.3.1 Samples are of high quality RNA

The quality of the RNA samples was consistently good although the quantity obtained from the tissues was relatively low, most notably in the CS16 FP samples (Table 4.2). Five of the eleven samples analysed with the Agilent Bioanalyser 2100 had RINs of 10.0, the highest value obtainable for this measurement. The sample with the lowest value (the E13.5 mouse pool sample) had a RIN of 9.6 (Table 4.2), which was still higher than that of the Ambion reference sample (the high quality standard used in this experiment) which has a RIN of 8.0 (Fig. 4.5A).

Table 4.2: Sample codes and corresponding sample numbers for all samples used in the microarray experiment. Where applicable the RIN number is given and the average absorbance ratio and average concentration is given for all samples.

Sample Code	Sample No.	RIN	Average A260/280 ± SD	Average Concentration (ng/µl) ± SD	Gel Picture Reference
E13.5 REF	E13.5	9.6	2.1±0	506±11	Fig 4.5A 11; B C
CS17 HP B1(T1)	MN45	-	2.0±0	146±0	Fig 4.5A 6
CS17 HP B2	MN40	-	2.1±0	130±3	Fig 4.5A 4
CS17 HP B3	MN44	-	2.1±0	112±3	Fig 4.5B 13
CS17 HP B4	MN31	-	2.0±0	135±3	Fig 4.5A 1
CS17 HP B1(T2)	MN45		2.0±0	146±0	Fig 4.5A 6
CS17 FP B1(T1)	MN45	10	2.0±0	71±2	Fig 4.5A 7
CS17 FP B2	MN40	-	2.2±0.1	54±1	Fig 4.5B 15
CS17 FP B3	MN44	10	2.1±0	77±1	Fig 4.5A 5
CS17 FP B4	MN31	-	2.0±0	76±1	Fig 4.5A 2
CS17 FP B1(T2)	MN45	10	2.0±0	71±2	Fig 4.5A 7
CS16 HP B1(T1)	MN60	-	2.0±0	90±1	Fig 4.5B 12
CS16 HP B2	MN39	10	2.2±0.1	75±3	Fig 4.5A 3
CS16 HP B3	MN49	-	2.1±0	77±1	Fig 4.5B 16
CS16 HP B4	MN59	10	2.0±0	101±4	Fig 4.5A 9
CS16 HP B1(T2)	MN60	-	2.0±0	90±1	Fig 4.5B 12
CS16 FP B1(T1)	MN60	9.8	2.0±0	72±0	Fig 4.5A 10
CS16 FP B2	MN39	-	2.2±0	40±1	Fig 4.5B 14
CS16 FP B3	MN49	10	2.0±0	45±0	Fig 4.5A 8
CS16 FP B4	MN59	-	2.0±0.1	49±2	Fig 4.5B 17
CS16 FP B1(T2)	MN60	9.8	2.0±0	72±0	Fig 4.5A 10

There was no evidence of degradation in any of the bat limb samples. Four of the eleven samples had DNA contamination which precluded the calculation of their RINs. These samples had similar electropherogram profiles to the samples with high RINs (Fig. 4.5A), and it was concluded that they were still very high quality samples. The six bat limb samples that were not analysed using the Agilent Bioanalyser 2100 were visualised using agarose gel electrophoresis. They were run against the high quality RNA from the E13.5 mouse pool sample. These samples also showed clear, bright bands corresponding to the 28S and 18S rRNA and there was no degradation, with all the samples showing a similar profile to that of the high quality E13.5 mouse pool sample (Fig. 4.5B).

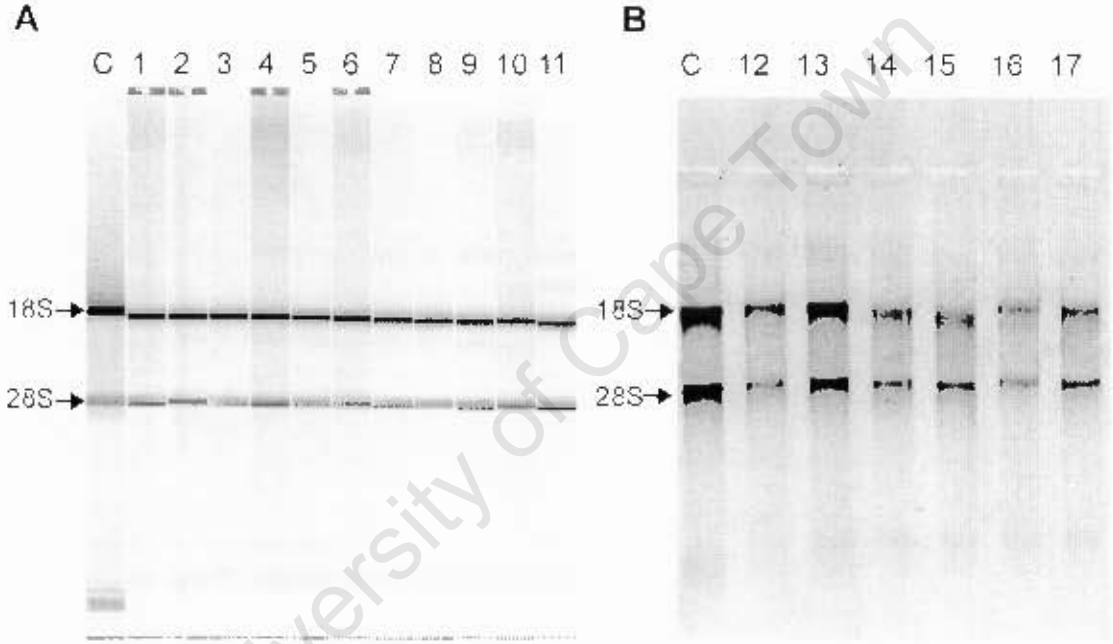


Fig. 4.5: Quality check of the integrity of the RNA samples was done through the use of (A) an electropherogram and (B) adenaturing agarose gel. The electropherogram shows that all samples were high quality RNA with distinct 28S and 18S bands occurring corresponding to those of the Ambion positive control (C). There was very little degradation which usually appears as smears below the 18S and 28S bands. Lanes 1,2 4 and 6 showed evidence of DNA contamination (indicated by the bar with a white dot above the lane) The denaturing agarose gel of the other samples indicated that these were also high quality RNA as two strong, clean bands occurred in each sample corresponding to those of the E13.5 mouse pooled ref sample which was used as a positive control (C).

The full absorbance profile of each sample was visualised between 220 nm and 320 nm and the 260/280 (A_{260}/A_{280}) absorbance ratio was calculated. This ratio gives an indication of the sample purity, indicating whether there is protein contamination. According to the Ambion Amino Allyl MessageAmp™ II aRNA Amplification Kit Instruction manual (Ambion, Austin, TX, USA) the A_{260}/A_{280} reading of a pure sample should fall within the range of 1.7 to 2.1 when the RNA is suspended in high quality water. The purity of the RNA samples was checked prior to experimentation and all the samples were pure with A_{260}/A_{280} values ranging from 2.0 to 2.2 (Table 4.2).

4.3.2 Analysis of amplification and labelling of experimental samples

The aRNA was assessed during the experiment by examining both its concentration and its size distribution on a gel after the amplification reaction. The aRNA yield after amplification and purification ranged from 83.3 µg to 110.9 µg (with an average of 102.1±1.4 µg). There was very little variation in aRNA concentration between samples both within and between batches, with the exception of the CS17 FP sample of Batch 4 (B4) which had a fairly low yield of 83.3 µg (Table 4.3).

Table 4.3: Total yield of aRNA (µg) for each sample and each batch after the IVT amplification reaction. Averages and standard errors (S.E.) are given in bold for both batches and samples.

Samples	B1	B2	B3	B4	Average±S.E.
E13.5 Mouse Pool	101.9	98.2	106.9	98.5	**101.4±2.0**
CS17 HP	110.9	95.0	102.6	103.1	**102.9±3.3**
CS17 FP	100.4	110.2	101.7	83.3	**98.9±5.7**
CS16 HP	109.1	101.6	102.8	107.3	**105.2±1.8**
CS16 FP	101.0	104.9	105.5	96.8	**102.1±2.0**
Average±S.E.	**104.7±2.2**	**102.0±2.7**	**103.9±1.0**	**97.8±4.6**	

The aRNA yield was not significantly different among the experimental samples nor among the batches (Anova: Two Factor without replication; samples: $F = 0.50$; $p > 0.05$; batches: $F = 1.13$; $p > 0.05$). All samples had aRNA yields above that of the expected yield of 64.1 µg for a reaction with 500 ng of starting material and an incubation time of 14 hrs. This difference may have been due to differences between the experimental RNA and the Ambion control RNA or it may have been due to the fact that the IVT reaction was maintained for 16 hrs (2 hrs longer than recommended). This yield indicates that the mRNA was amplified from 3 330-fold to 4 435-fold (an average of 4 083±55-fold) assuming that the original mRNA component of the RNA sample was 5% (Ding et al., 2006). After purification all of the samples had an A_{260}/A_{280} of 2.0. The size distribution of the all aRNA was similar both within and between batches appearing to range from about 250 to 3 000 nt, with highest density occurring below the 18S RNA band (Fig. 4.6). Very little signal was seen from the 5 000 to 2 000 nt transcripts, which has been attributed to loading the minimum amount of aRNA that was recommended (1 µg) resulting in a gel with a relatively weak signal and a low resolution of low abundance transcripts. The aRNA size distribution was similar to that reported in previous studies (Ding et al., 2006).

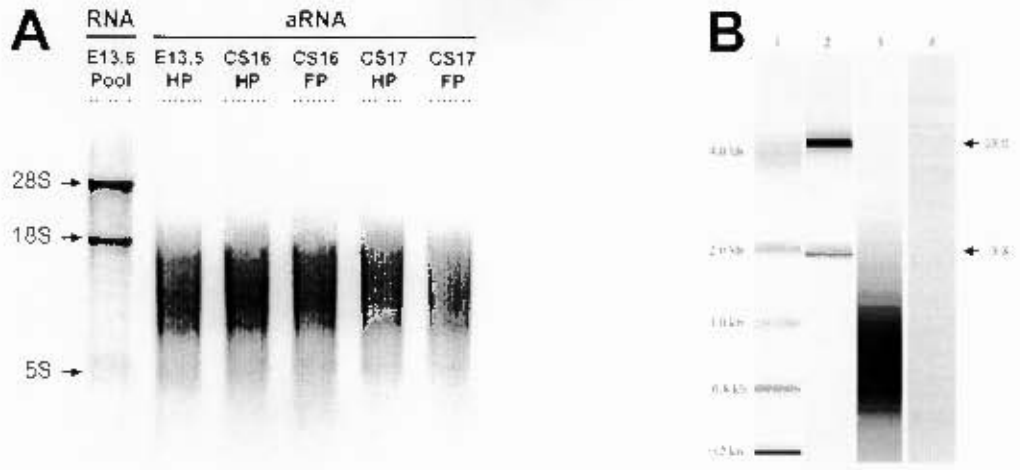


Fig. 4.6: Denaturing agarose gel showing (A) the size distribution of the aRNA samples for B1. The aRNA size distributions should appear as a smear from 4 000 nt (adjacent to the 28S RNA band) to 200 nt with the highest density occurring around 1 400 nt (below the 18S band). The distribution of aRNA is comparable to that reported in previous studies. (B) A virtual gel created from Bioanalyser results in which 1 is a molecular marker, 2 is a RNA sample, 3 is a aRNA sample and 4 is a negative control (Figure B taken from Ding et al. (2006)).

The percentage dye incorporation was measured for all samples and showed a significant difference between each group and across all of the batches (Anova: Two Factor without replication; samples: $F = 34.98$; $p < 0.01$; batches: $F = 11.82$; $p < 0.01$). There were no significant differences among the test sample groups (Cy3 labelled aRNA sample groups) when the reference sample group (Cy5 labelled aRNA sample groups) was excluded, although there was still a significant difference among the batches in this analysis (Anova: Two Factor without replication; samples: $F = 1.81$; $p > 0.05$; batches: $F = 14.5$; $p < 0.01$). This can be attributed to the relatively higher dye incorporation rates of the first experiment (B1-T1) which were on average higher that those of all the other batches (Table 4.4). There was no significant difference in the Cy3 (green) dye incorporation rates of experiments B1-T2 to B4 (Anova: Single Factor; $F = 1.0$; $p > 0.05$).

Table 4.4: Dye incorporation for each aRNA sample and each batch after the labelling reaction. Averages and standard errors (S.E.) are given in bold for both batches and samples.

Samples	B1-T1	B1-T2	B2	B3	B4	Average±S.E.
E13.5 Mouse Pool	47	36	39	46	36	**41±2.4**
CS17 HP	74	60	65	59	60	**64±2.8**
CS17 FP	76	60	61	61	60	**64±3.1**
CS16 HP	72	51	65	54	60	**60±3.8**
CS16 FP	68	60	56	57	60	**60±22.2**
Average±S.E.	**73±1.5**	**58±2**	**62±2**	**58±1.4**	**60±0**	

The difference in the dye incorporation between the reference and the test sample groups is a result of the different dye incorporation rates of the two fluorescent dyes with the Cy5 (red) dye incorporating less efficiently (41±2 dye molecules per 1000 nt) than the Cy3 (green) dye (62±2 dye molecules per 1 000 nt). However, these readings were close to the expected range of 30-60 dye molecules per 1 000 nt). The reason for the difference between the dye incorporation rates of B1-T1 and that of subsequent batches is unknown. The concentration of each dye in the samples to be loaded was also determined and all the samples were beneath or close to the recommended maximum concentration of 0.8 pmol/µl (Table 4.5).

Table 4.5: Dye concentration (pmol/µl) for each hybridisation solution sample and each batch. Averages and standard errors (S.E.) are given in bold for both batches and samples.

Samples	B1-T1	B1-T2	B2	B3	B4	Average±S.E.
E13.5 Mouse Pool	0.52	0.49	0.42	0.63	0.55	**0.52±0.03**
CS17 HP	0.77	0.67	0.68	0.77	0.55	**0.69±0.04**
CS17 FP	0.73	0.60	0.59	0.80	0.65	**0.67±0.04**
CS16 HP	0.79	0.57	0.69	0.70	0.65	**0.68±0.03**
CS16 FP	0.81	0.79	0.64	0.73	0.61	**0.72±0.04**
Average±S.E.	**0.77±0.02**	**0.62±0.05**	**0.61±0.05**	**0.73±0.03**	**0.60±0.02**	

Dye concentrations were kept below this value to prevent excessive background fluorescence occurring on the slides after hybridisation. There was no significant difference in the dye concentration between the different samples and the different batches (Anova: Two Factor without replication; samples: $F = 0.48$; $p > 0.05$; batches: $F = 0.92$; $p > 0.05$).

4.3.3 Batch excluded due to wash effects

Microarray images were processed to capture the signal of each spot based on the expected location of the spot and the contrast between the spot fluorescence and that of the background. Microarrays from different batches did not exhibit any differences in the average number of spots that had an array signal, B2, however, did have a much lower incidence of spots flagged as bad (Table 4.6). Two arrays stood out due to their remarkably low signal, these were CS17 FP B1-T2 (8%) and CS17 FP B2 (17%) with the majority of their spots flagged as absent. This loss of signal was attributed to photo bleaching that was noted during the scanning process. On average, over all the arrays, 35% of the spots showed a slide signal, 50% were marked as absent and 15% were flagged as bad.

Table 4.6: Percentage of good and flagged features on each array including the average percentage of flagged features for each batch

Batches	Sample	Percentage (%)			Average batch percentage (%)	
		Prosont signal	Flagged bad	Flagged absent	Present signal	Flagged bad and absent
B1-T1	16 HP	46	19	35	34±9	66±9
	17 HP	32	22	46		
	16 FP	35	22	43		
	17 FP	23	15	62		
B1-T2	16 HP	64	24	12	39±24	61±24
	17 HP	51	22	27		
	16 FP	33	13	54		
	17 FP	9	4	87		
B2	16 HP	36	18	46	31±10	69±9
	17 HP	35	3	62		
	16 FP	37	5	59		
	17 FP	17	6	77		
B3	16 HP	38	13	49	40±9	60±10
	17 HP	34	19	48		
	16 FP	54	17	29		
	17 FP	35	12	54		
B4	16 HP	29	11	60	33±4	67±5
	17 HP	35	15	49		
	16 FP	29	12	59		
	17 FP	37	21	41		

Visualisation of slide signal images indicated that B1-T1 arrays had a high incidence of wash artefacts. These occurred in the green channel (Cy3) and were noted around the perimeter of the array, showing a high degree of spatial bias in comparison to B1-T2 arrays (Fig. 4.7). A conservative approach was taken and this poor quality batch was removed rather than merging the two technical replicates.

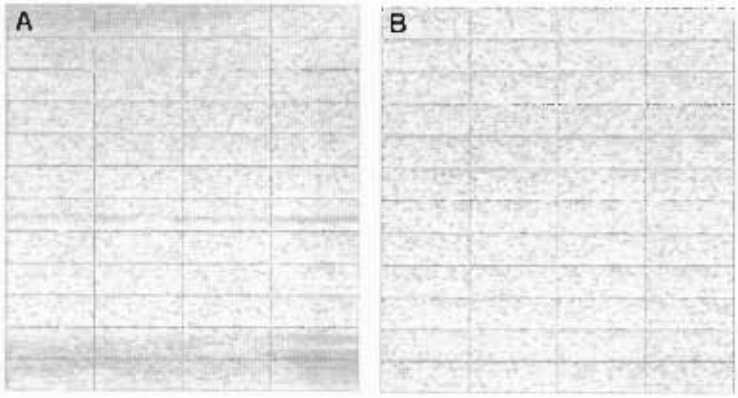


Fig. 4.7: Comparison of spot signal intensity for the Cy3 (green) channel of (A) technical repeat 1 given by array CS16 HP B1-T1 and (B) technical repeat 2 given by array CS16 HP B1-T2. Spatial effects were apparent in all arrays of technical repeat 1, while those of technical repeat 2 were considered to be of a higher quality.

4.3.4 Impact of transformation steps on the data set

It was not technically feasible to process twenty arrays in a single experimental session; instead they were separated into batches of four arrays each. The data transformation pipeline involved a multi step normalisation and pre-processing pathway with a variety of checks to validate the chosen methods (Fig. 4.9).

An examination of arrays prior to normalisation showed the presence of batch effects, this occurs when the arrays of a batch are more similar to one another than to those of its corresponding biological repeat of another batch. The 48 tiling spots (dye markers) spotted on each array were found to have a very strong signal (average A value = 12.7±0.7) with a strong bias towards the Cy3 channel (average Cy3 log intensity = 15.3±0.9; average Cy5 log intensity = 12.7±0.7) resulting in heavily skewed log ratios at high intensities (average M value = -5.2±1.1). These outliers were therefore removed prior to normalisation. Visualisation of the A value distributions of the estimated background, null spots, buffer spots and negative controls of all the arrays prior to normalisation indicated that the average estimated background signal (5.8±0.6) was comparable to the signal from the null spots (5.8±0.7). It was similar to that of the negative controls (6.1±0.6), but lower than that of the buffer spots (6.3±0.7, Fig. 4.8).

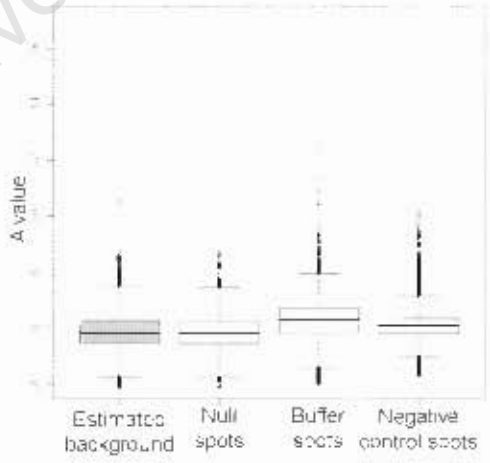
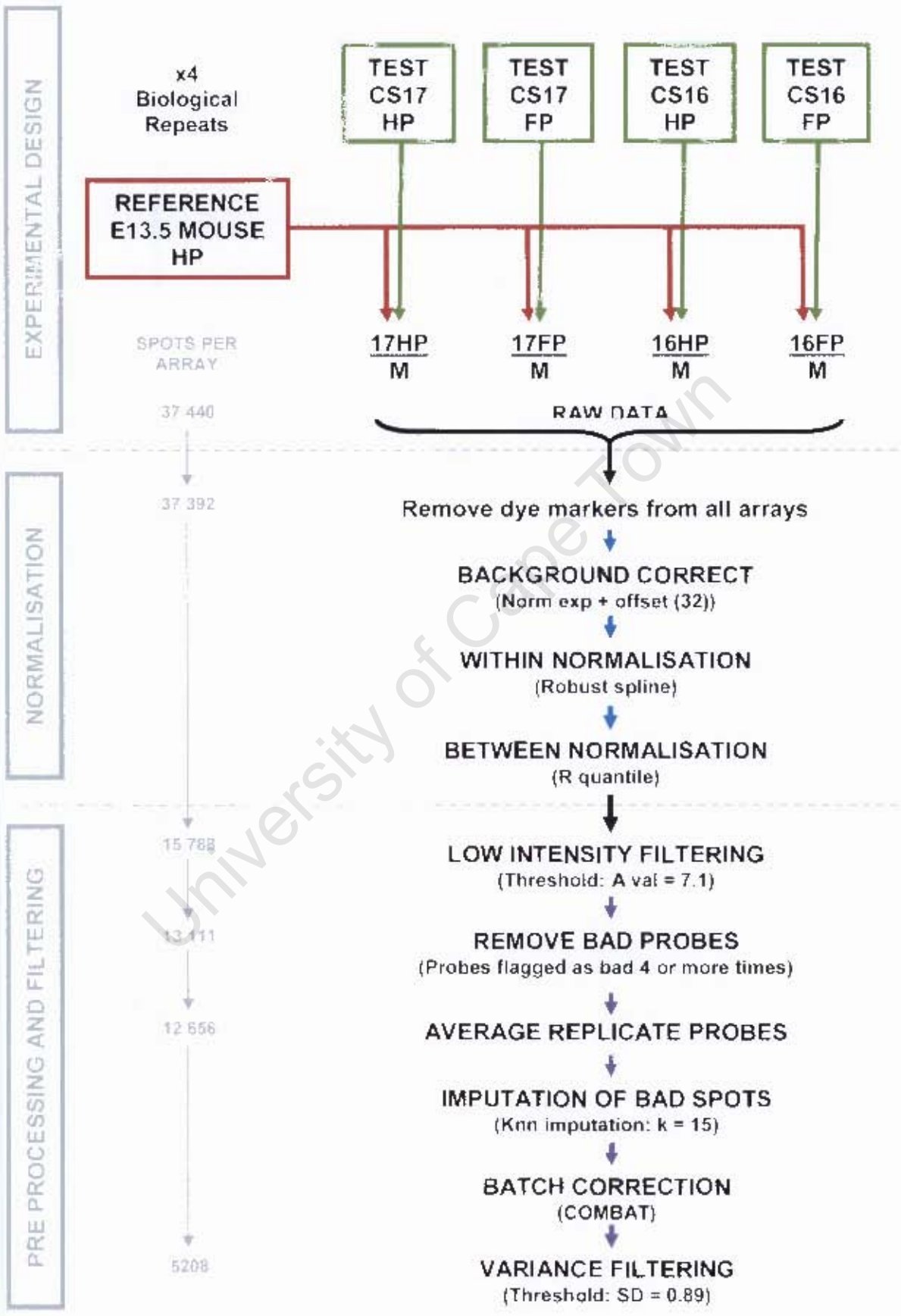


Fig. 4.8: Box plot of A value distribution of the estimated background (grey), and the spots used as low intensity signal estimators (white). The estimated background has a very similar distribution as that of the null spots indicating that it is an accurate parameter to use when correction for background effects.

This indicated that the median estimated background spot parameter was a good indicator of low or no signal, and could be used as a suitable parameter for correcting for the biases that are due to this non-specific binding. This was done using the *normexp + offset* method, which has been shown to outperform conventional methods in a comparative study of background correction methods (Ritchie et al., 2007).

Thirty-two was chosen as the offset value as it resulted in an average distribution of the null and negative control spots that most closely resembled the original distribution, whilst still correcting for batch specific differences. The raw data had a dramatic dye bias, with the green channel having greater overall intensity (7.37±1.29) than the red (6.34±1.57) this resulted in the log ratios being on average negative (-1.03±0.96). The strong batch effects were apparent when looking at the distributions of the log ratios (Fig. 4.10B) and the hierarchal clustering tree (Fig. 4.11C). Background correction had a remarkable effect on the data, reducing batch effects (Fig. 4.10E, Fig. 4.11E & F) and decreasing the variance between the arrays (Fig. 4.11A & D). This occurred as a result of the individual channel correction which reduced the dye bias between the red and green channels at lower intensity levels (Fig. 4.10D). Batch effects that were a result of scanner settings and wash artefacts had a greater impact on low intensity values whose log ratios are less stable. Background correction should selectively adjust these susceptible values without introducing bias due to overcorrection. The addition of an offset to the background corrected data contributed to the standardisation of the channels at lower intensities as it made these log ratios tend towards zero. Dye bias was still apparent at higher intensity levels (Fig. 4.10D) but as lower intensity values made up the majority of the signal on the array (Fig. 4.10C), this correction had the effect of bringing the overall distribution of the M values (the logged ratio of the red and the green channel) closer to zero (-0.48±0.70).

The distribution of the A values remains fairly unchanged however it can be seen that the addition of the offset of thirty-two (corresponding to an A value of five) has the effect of standardising the lower intensity values across arrays. Robust spline normalisation corrected intensity dependant and special bias that was found on the arrays.



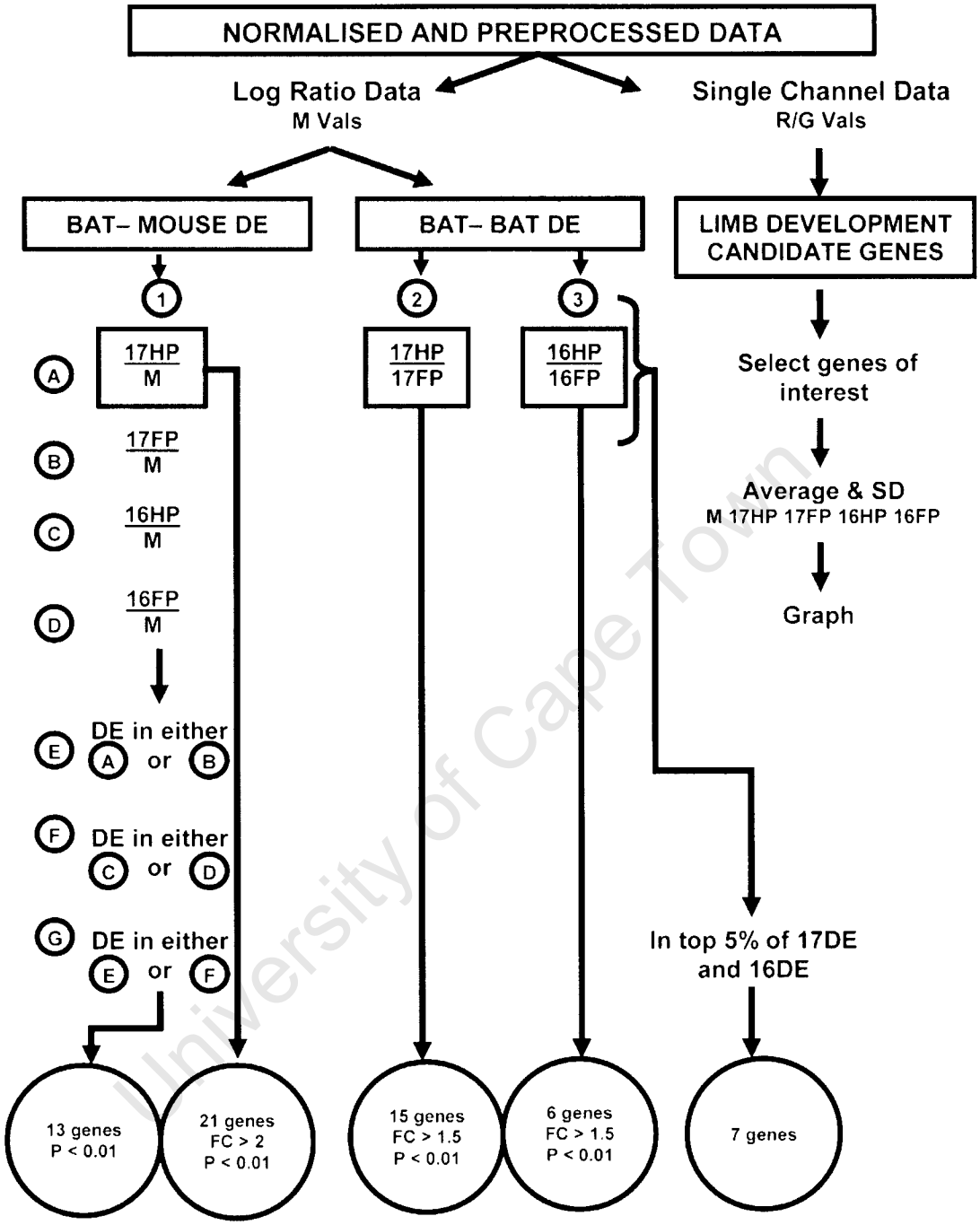


Fig. 4.9: Flow diagram of experimental design, normalisation, preprocessing and data analysis steps performed in this experiment. The number of spots remaining on each array is given on the left, through out the preprocessing pipeline and the number of genes (including the statistical thresholds used) is given for each differential expression (DE) analysis.

This method balanced the red and the green channel (Fig. 4.10G), centring the M value distribution around zero (Fig. 4.10H). This was done without forcing the data to take on the same distributions. Batch effects are still apparent with B1 having a very small inter-quartile range while B3 had a larger one.

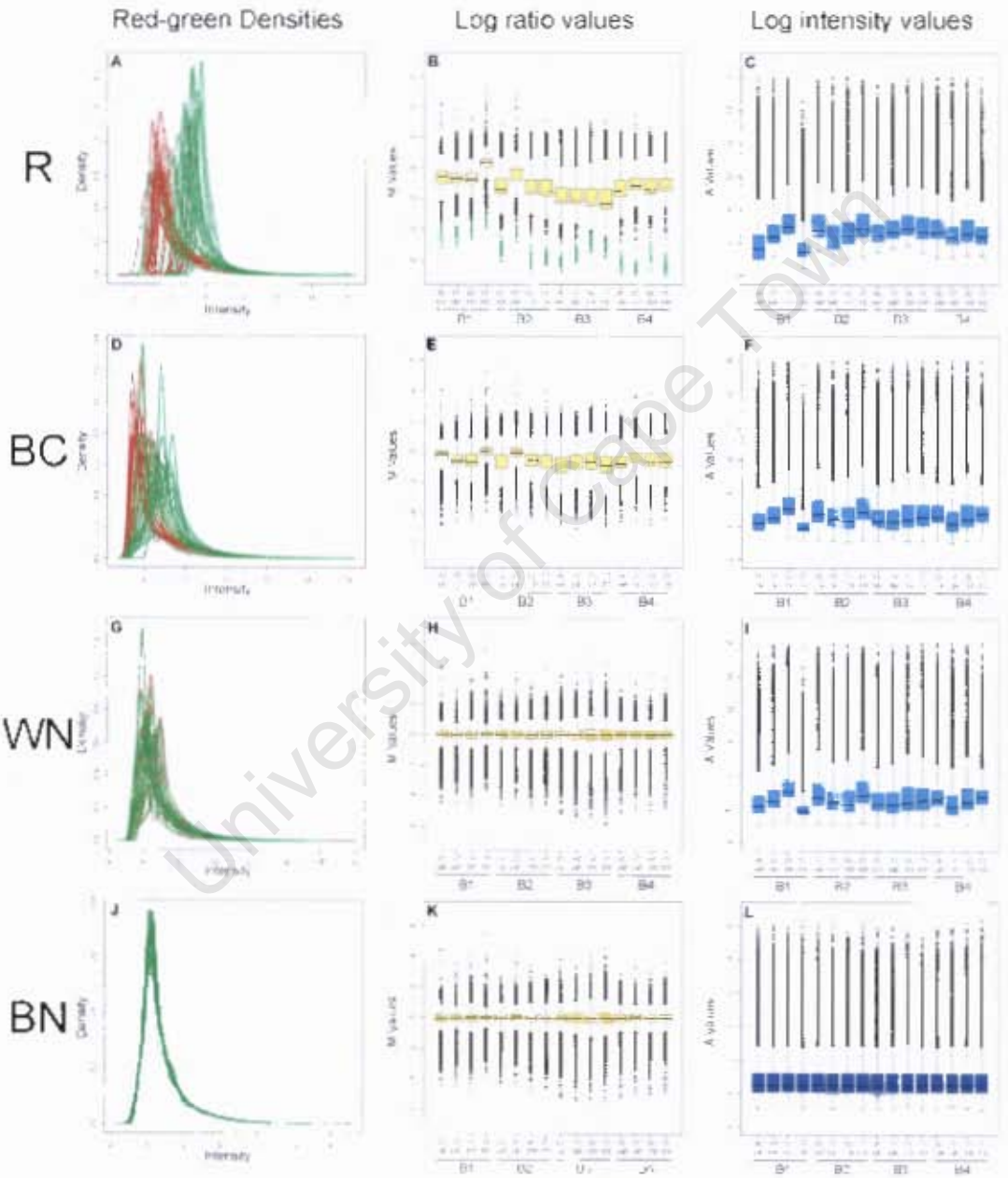


Fig. 4.10: Visualisation of data after each step in the normalisation pipeline. Raw (R), background corrected (BC), within normalised (WN) and between normalised (BN) data distributions become more similar after each step. Density distribution of the both the green and red signal intensities are given. Un-normalised log ratios (M values) are shown in light yellow and normalised log ratios are shown in yellow. Un-normalised log intensities (A values) are shown in light blue and normalised log intensities are shown in blue.

4.3.5 Preprocessing and filtering refines data and reduces batch effects

Normalised data were processed to remove low intensity values, correct for batch effects and to impute missing values (Fig. 4.9). It was assumed that the signal of the majority of transcripts on each whole genome array would remain unchanged across all arrays, based on this assumption there should be a good correlation of the intensity (A values) between all of these arrays. The correlation between normalised arrays improved as lower quantiles of intensity values were sequentially removed (Fig. 4.13A), with the maximum correlation (r^2 = 0.902) occurring when the lower 90% of each array was removed. This indicates that variability at lower intensities was responsible for the differences between the arrays. This is not an artefact of spot number reduction because when this operation is done for the higher quantiles of intensity values it has the opposite effect, making arrays more dissimilar as more values are removed (Fig. 4.13A). This indicates that low intensity filtering would be a suitable strategy to reduce noise among arrays, with the greatest improvement in correlation occurring over the first 30% of lower intensities removed. To do this a threshold level, below which signal was considered extremely low, was calculated from the signal of the normalised null (5.9±0.5), buffer (6.3±0.5), and negative control (6.1±0.5) spots.

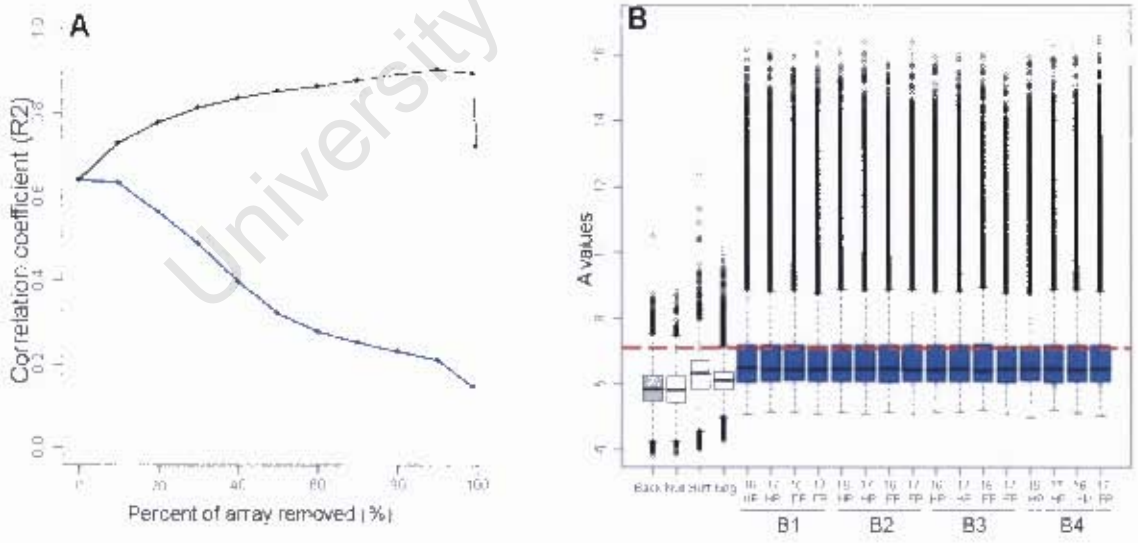


Fig. 4.13: (A) Removal of the lower quantiles of data (black line) improves the overall correlation between the data. This relationship is not found when the higher quantiles are removed (blue line) indicating that low intensity filtering should be performed. (B) Boxplots of A value distribution of the estimated background (Back) in grey, the null spots (Null), the buffer spots (Buff) and the negative control spots (Neg) in white and all the spots of each array in blue. The filter threshold below which spots are considered as having a low A value is shown as a dashed red line. The majority of the spot A values spots for fell below this threshold.

This threshold was calculated as the average maximum value (excluding outliers) for these spot groups (7.1±0.3). On average 73% of the array spots fell below this threshold (Fig. 4.13B) indicating that the arrays had an overall low signal. Probes with spot signals that consistently fell below this threshold across all arrays were removed. This filtering step excluded 58% of the probes (21 604 spots on each array), leaving 15 778 probes for further processing. Twenty two percent (8 277 spots on each array) of probes were flagged or weighted as bad more than four times across all arrays. When these were excluded an additional 7% of each array (2 677 spots on each array) was removed. After these two filtering steps 35% of the probes (13 111 spots on each array) were left for further pre-processing. This included 65% of the null spots, 22% of the buffer spots and 8% of the negative controls. This indicates that, though the filtering threshold for low intensity (an A value of 7.1) is quite high relative to the average array signal, the filtering itself is quite a conservative and selective one. It allowed probes with low intensity, variable signal to be retained, while probes with a consistently low signal were excluded. This filtering step did not create a bias between arrays. The density distribution of the red and green channels, the M values and the A values remained comparable across arrays (Fig. 4.14A, B & C). However, it did appear to increase the batch effect found among arrays (Fig. 4.15B & C). Arrays varied in the number of weighted values (2–12%) that remained post filtering, in total 6% of all the spots over all arrays (11 699 spots) were replaced with missing values. Averaging probe replicates reduced the number of probes in the analysis from 13 111 to 12 656 probes.

Values that were weighted as bad were removed and imputed using KNN imputation. It was shown that the NRMS error in a test data set (with randomly generated missing values) was minimised when k was set at between 15 and 20. The value of k was set at 18 as this value was found to be the most robust value when minimising the NRMS error of the test dataset.

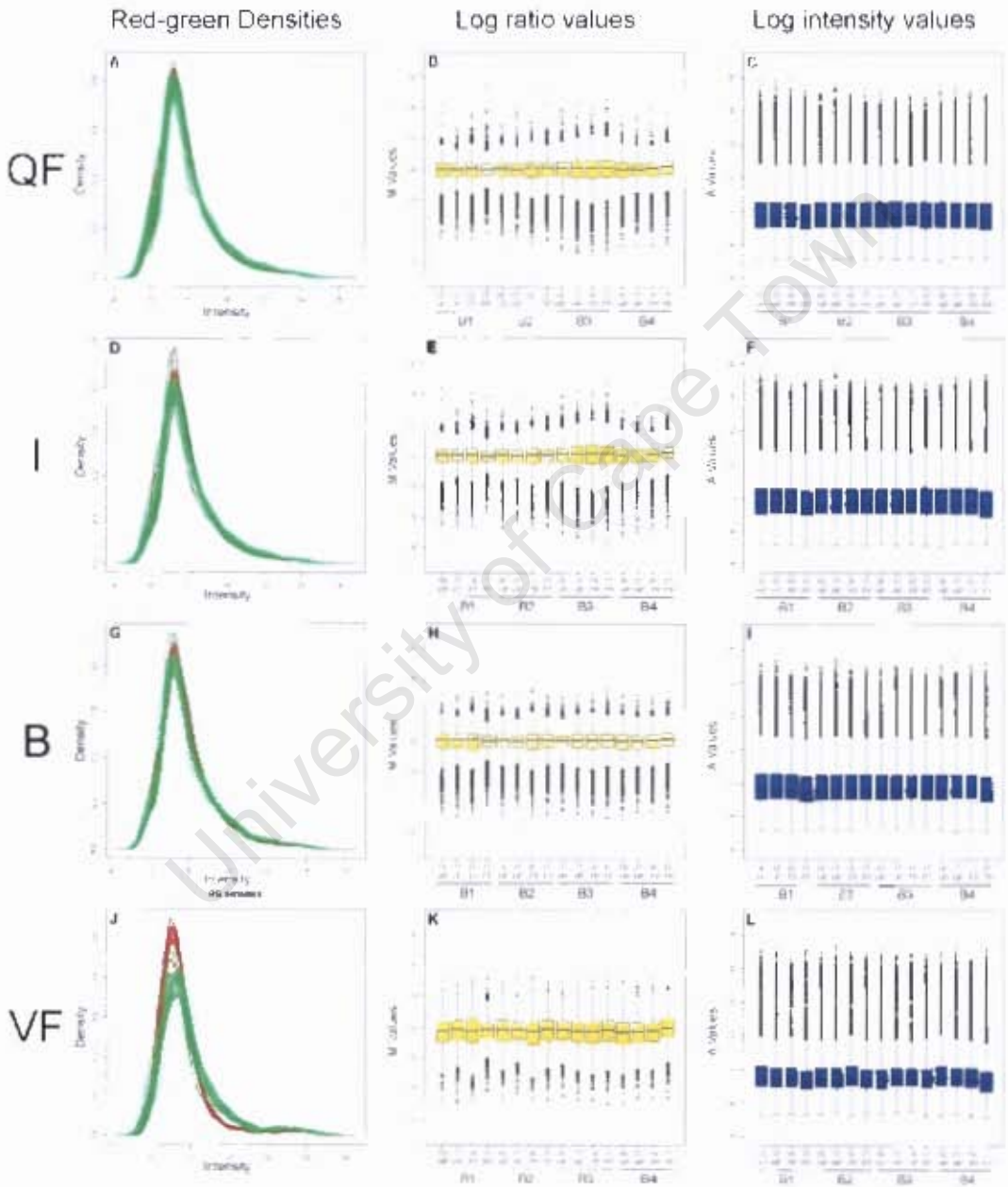


Fig. 4.14: Visualisation of data after each step in the pre-processing pipeline. The similarities of the data distributions amongst arrays was maintained throughout quality filtering (QF), imputation (I), batch correction (B) and variance filtering (VF). Density distribution of the both the green and red signal intensities are given. Normalised log ratios are shown in yellow and normalised log intensities are shown in blue.

4.3.6 Preliminary analysis shows close association between groups

A comparison of the log transformed signal between the different limb condition groups showed that there were much larger differences between the cross species comparison groups (Fig. 4.18A & B) than the between the within species comparison groups (Fig. 4.18C & D). A comparison of the average mouse hand plate signal and the average CS17 bat hand plate signal showed that though there was a high degree of correlation between the two samples ($r = 0.887$, $p < 0.01$) there were also a large number of probes which exhibited large (>2) fold changes (Fig. 4.18A). In this comparison 3.67% of the probe set of 37 392 probes were more than two fold higher in the bat hand plate while only 0.94% were found to be over two fold higher in the mouse hand plate (Fig. 4.18A).

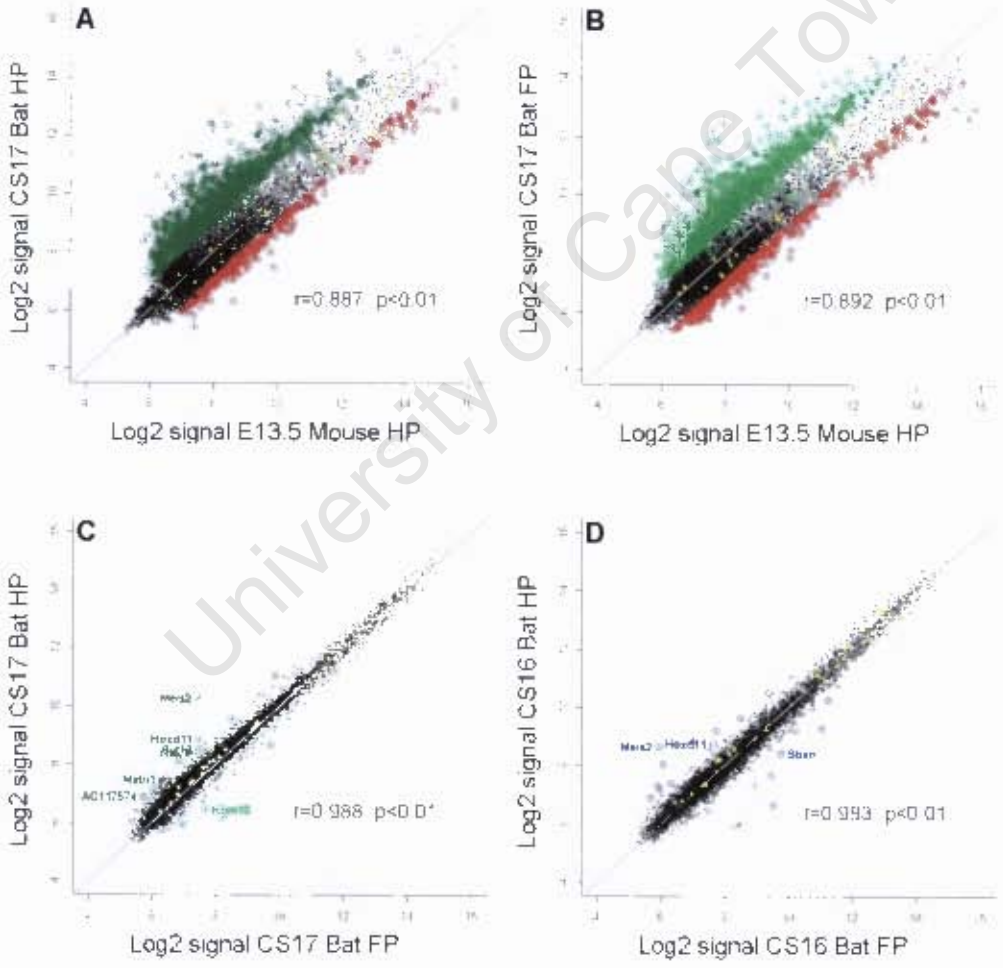


Fig. 4.18: Scatterplots of average transcript signal similarity for normalised and pre-processed data. Comparisons were made between the signal from (A) the CS17 bat HP (dark green) and the mouse (red); (B) the CS17 bat FP (light green) and the mouse (red); (C) the bat CS17 HP and the FP and (D) the bat CS16 HP (dark blue) and FP (light blue). All comparisons were highly correlated however greater signal differences were in the bat-mouse comparisons than in the bat-bat comparisons

In this comparison 3.67% of the probe set of 37 392 probes were more than two fold higher in the bat hand plate while only 0.94% were found to be over two fold higher in the mouse hand plate (Fig. 4.18A). There was also a very loose distribution of the OpArray control (housekeeping genes) with many exhibiting high fold changes between the bat and the mouse samples. This bias in fold change between the two groups was repeated in the comparison between the mouse hand plate and the CS17 bat foot plate which showed a comparable correlation ($r = 0.892$, $p < 0.01$) and fold changes (3.21% over two fold higher in bat foot plate while 1.29% were higher in mouse hand plate (Fig. 4.18B). A comparison of the probes with high fold changes in the inter-species comparison showed that 80% of these probes were similarly regulated across all bat limb conditions.

This contrasted greatly with the intraspecies comparisons which showed extraordinary conservation between their gene-wise signals. A comparison between the CS17 hand and foot plates showed very high correlation ($r = 0.988$, $p < 0.01$) and a very low number of probes with high fold change differences (0.04% over two fold higher in the hand plate while 0.01% were higher the foot plate) (Fig. 4.18C). In these comparisons the OpArray control genes were tightly distributed and did not exhibit high fold changes. These findings were similar to the comparison between the CS16 hand and foot plates groups, which exhibited high correlation ($r = 0.983$, $p = 0.01$) and a low number of probes with high fold changes (0.03% over two fold higher in the hand plate while 0.03% were higher in the foot plate (Fig. 4.18D). This pattern of association between groups was consistent among all group comparisons (data not shown).

4.3.7 Dataset reveals biologically valid information

The reliability of the array signal in detecting differential expression was tested by looking at the expression of specific genes that have been validated as being differentially expressed between the hand and the foot plate of mouse E11.5 and E12.5 limbs (Margulies et al., 2001; Shou et al., 2005). The genes *Tbx5*, *Hhip*, *Hoxd11*, and *Hoxd10* have been shown to be significantly up-regulated in the mouse E12.5 hand plate while *Pitx1*, *Hoxa11*, *Tbx4* and *Hoxc10* were found to be significantly up-regulated in the mouse foot plate at this stage (Shou et al., 2005). The expression of these genes was examined in the CS16 bat hand and foot plate, a comparatively similar bat stage to that of the mouse at E12.5 (Table 4.7). Though the relative expression of these genes in the bat shows a similar overall trend to that in the mouse in terms of fold changes, only one gene (*Hoxd11*) was significantly differentially expressed.

The important fore- and hindlimb specific genes, *Tbx4*, *Tbx5* and *Pitx1*, had very low signals on the arrays. In the case of *Pitx1* this can be attributed to the low similarity of the probe to the bat sequence (59%) causing there to be low to no signal for the bat channel (Table 4.7). However sequence information was not available for bat *Tbx4* and *Tbx5* (Table 4.7). The three probes that show high similarity of the bat sequence (*Hoxd11*, *Hoxd10* and *Hoxc10*) also show the same trend as that of the mouse in terms of fold change between the hand and foot plate.

Table 4.7: The signal of CS16 bat hand plate (HP) and foot plates (FP) for genes that are known to be differentially expressed in the developing hand plate and foot plates of the mouse stage E12.5 are shown. The signal is the mean transcript signal from normalised single channel data of four biological repeats. Fold changes (FC) between the signal of the hand and the foot plates are given for both the bat and the mouse (Shou et al., 2005). * indicates that a gene is significantly differentially expressed between the two conditions ($p < 0.05$). NS indicates that there was no sequence information for the bat gene over the potential probe hybridization site.

OPERON ID (Accession No.)	Gene Name	Limb specific	% Similarity	Signal Bat HP (SD) †	Signal Bat FP (SD) †	FC Bat HP/FP	FC Mouse HP/FP
M400000286 (NM_011537)	*Tbx5*	HP	NS	122 (±8)	116 (±7)	1.1	16*
M200008617 (NM_020259)	*Hhip*	HP	NS	244 (±84)	202 (±86)	1.2	3.3*
M200002041 (NM_008273)	*Hoxd11*	HP	99	424 (±87)	199 (±26)	2.1*	1.4*
M200004888 (NM_013554)	*Hoxd10*	HP	97	237 (±33)	168 (±0.9)	1.4	1.4*
M300003311 (NM_010462)	*Hoxc10*	FP	90	110 (±18)	142 (±51)	-1.3	-2.2*
M300000006 (NM_011536)	*Tbx4*	FP	NS	68 (±15)	64 (±3)	1.1	-2.8*
M300011421 (NM_010450)	*Hoxa11*	FP	NS	1003 (±151)	1201 (±41)	-1.2	-3.4*
M200001984 (NM_011097)	*Pitx1*	FP	59	91 (±16)	105 (±33)	-1.2	-4.6*

A FatiScan comparative functional analysis was done between the genes that exhibited strong signal (>300) in the mouse and the genes that exhibited a strong signal (>300) in the bat (both in the HP and in the FP). No significant differences were found in the functional annotations of the mouse specific gene list and that of either bat limb conditions (adjusted $p < 0.05$). This indicates that the functional profile of the genes which exhibit a signal in the mouse are not significantly different to that of the equivalent bat HP and FP.

4.3.8 Candidate genes selected through differential expression analysis

The DEDS statistic was used to rank probes in order of their differential expression, the significance of these differences was then determined by calculating the BH-FDR (q) of each probe (hereafter referred to as genes). Three comparisons of interest were explored to screen for differentially expressed (DE) genes (Fig. 4.9). The first was the inter-specific comparison between the mouse E13.5 hand plate and the bat CS17 HP and FP; the second was the intra-species comparison between the CS17 HP and FP and the third was the intra-species comparison between the CS16 HP and FP.

i Inter-species comparison of (Mouse HP vs. Bat HP and FP)

Three hundred and fifty genes were significantly differentially expressed between the mouse HP and the bat CS17 HP and three hundred and twelve genes between the mouse HP and the bat CS17 FP ($q < 0.01$). However, the majority of the genes (247) were the same across both lists (Fig. 4.19) indicating that these genes may have been identified due to species-specific differences in transcript abundance or hybridisation ability of the individual transcripts, rather than limb morphology differences between the bat and the mouse. These redundant genes were removed by filtering the HP and FP lists against one another for both CS17 and CS16 (Fig. 4.19). This left 168 candidate genes that were differentially expressed between the bat limb types and the mouse hand plate at CS17.

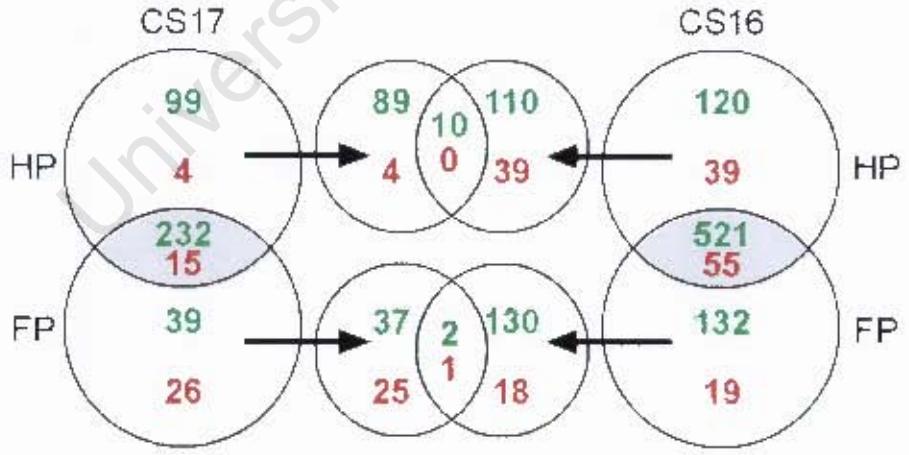


Fig. 4.19: Venn diagram showing how the different limb conditions were used as biological filters to identify robust candidate genes. Genes that were overexpressed in the bat limb compared to the mouse limb are shown in green while those that were underexpressed are shown in red. Greyed out areas indicate genes that were excluded from the gene lists as they were found to be significantly differentially expressed in both the hand and the foot plate of the bat. These limb specific DE genes were compared to lists obtained from the CS16 bat comparison. In this way thirteen genes were isolated that were robustly expressed either in the bat HP or FP as compared to the mouse but not in both.

Table 4.8: Summary of genes differentially expressed at a level 2 fold or higher between the mouse hand plate and CS17 bat hand and foot plates. Genes in bold indicate those that are similarly regulated in the CS16 bat autopods. † are given when the gene associated with the probe was annotated through blast analysis. The gene ontology (GO) of the molecular and biological (in brackets) function is also shown. The percentage similarity (%) between the sequence of the mouse OPERON probe and the homologous bat gene is shown, with NS indicating that no sequence was available for the bat for that gene. The fold change (FC) between the CS17 HP and the mouse hand plate and the CS17 FP and the mouse foot plate is also given. * signifies whether this change is statistically significant ($q < 0.01$). Genes are ordered in terms of their fold changes between the CS17 HP and the mouse hand plate

OPERON ID (Accession No.)	Gene Name	GO: Molecular (Biological)	%	FC CS17 HP	FC CS17 FP
M400017713 (A830011L22rik)	Meis homeobox 2 (Meis2)†	Transcription factor activity (Regulation of transcription)	NS	9.91*	2.19
M300005520 (NM_030725)	synaptotagmin XIII (Syt13)	Transporter activity (Vesicle-mediated transport)	45	3.95*	1.95
M400014732 (AC108415)	Nipped-B homolog (Drosophila) (Nipbl)†	Binding (Cell cycle)	NS	3.44*	2.47
M400002772 (AC093175)	predicted gene, EG668529†		NS	3.34*	2.62
M400001196 (NM_010063)	dynein cytoplasmic 1 intermediate chain 1 (Dync1i1)	Microtubule motor activity (Microtubule-based movement)	NS	3.33*	2.31
M200015556 (NM_027165)	Cyclin-dependent kinase 3 (Cdk3)†	Nucleotide binding; protein kinase activity (Cell cycle, Cell division)	64	3.26*	1.84
M400002554 (XM_984784.1)	PR domain containing 2, with ZNF domain (Prdm2)	Nucleic acid binding; transcription regulator activity	83	3.22*	2.17
M200000361 (NM_008858)	protein kinase D1 (Prkd1)†	Protein kinase activity; nucleotide binding (Intracellular signalling cascade)	NS	3.21*	2.79
M300002857 (NM_013500)	Hyaluronan and proteoglycan link protein 1 (Hapln1)	Hyaluronic acid binding (Cell adhesion)	NS	3.13*	1.91
M200005039 (XM_980284.1)	cDNA sequence BC010304		NS	3.08*	2.66*
M400005171 (XM_126489)	Forkhead box K2 (Foxk2)	Transcription factor activity (Regulation of transcription)	89	3.06*	2.43
M300013935 (NM_175506)	a disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif, 19 (Adamts19)	Peptidase activity (Proteolysis; integrin-mediated signalling pathway)	87	2.91*	2.01
M300000971 (NM_008687)	nuclear factor I/B (Nfib)	Transcription factor activity (Neg. regulation of cell proliferation)	93	2.89*	2.34
M200006565 (NM_175641)	latent transforming growth factor beta binding protein 4 (Ltbp4)	growth factor binding (Regulation of cell growth and differentiation)	87	2.87*	2.43
M400005372 (AC125323)	RIKEN cDNA C920008N22 gene†		NS	2.86*	2.08
M200014991 (NM_001002012)	heat shock protein 2 (Hspa2)	Protein binding (Protein folding; response to unfolded protein)	93	2.81*	2.62

Table 4.8: (Continued)

OPERON ID (Accession No.)	Gene Name	GO: Molecular (Biological)	%	FC CS17 HP	FC CS17 FP
M400015787 (AC099702)	AC099702		NS	2.80*	2.44
M400013097 (1810038L18rik)	similar to Dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2 (LOC100044376)†	Nucleotide binding, Protein kinase activity (Protein amino acid phosphorylation)	NS	2.73*	2.26
M200002276 (NM_080556)	transmembrane 9 superfamily member 2 (Tm9sf2)	Transporter activity (Transport)	NS	2.70*	2.22
M400009657 (NM_178660.2)	RNA binding motif, single stranded interacting protein (Rbms3)	Nucleotide binding; RNA binding	NS	2.67*	2.29
M300003443 (NM_026439)	coiled-coil domain containing 80 (Ccdc80)		NS	2.64*	2.80
M400013975 (AC110038)	RIKEN cDNA 1700040K01 gene*		94	2.64*	2.45
M400005172 (XM_357872)	RIKEN cDNA 2900075B16 gene	Calcium ion binding	NS	2.61*	2.08
M300012366 (NM_027707)	Nipped-B homolog (Drosophila) (Nipbl)	Binding (Cell cycle)	93	2.57*	1.99
M400010530 (AC140268)	Mus musculus BAC clone RP23-321K4 from chromosome 7†		NS	2.37	2.96*
M200001308 (NM_008778)	p21 (CDKN1A)-activated kinase 3 (Pak3)	Nucleotide binding, Protein kinase activity (Multicellular organismal development)	NS	2.21	2.59*
M200007701 (NM_011132)	polymerase (DNA directed), epsilon (Pole)	DNA-directed DNA polymerase activity, transferase activity (DNA replication)	NS	-1.75	-2.47*
M400018107 (AC117574)	RIKEN cDNA 4930428O21 gene†		NS	-1.92	-3.45*
M400003965 (NM_001033530.1)	expressed sequence AW146154	Nucleic acid binding (Regulation of transcription)	66	-1.99	-2.61*
M200002968 (NM_009183)	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4 (St8sia4)	Transferase activity (Protein amino acid glycosylation)	NS	-2.11	-2.89*
M400001512 (NM_177712)	RIKEN cDNA C330011K17 gene	Nucleic acid binding (Regulation of transcription)	NS	-2.11	-2.99*
M400012224 (NM_172427)	RIKEN cDNA 2310067E19 gene†		NS	-2.20	-3.07*
M300016983 (NM_198425)	EP300 interacting inhibitor of differentiation 2 (Eid2)	Protein binding (regulation of cell proliferation)	NS	-2.31	-3.13*
M400010868 (NM_008907)	peptidylprolyl isomerase A (Ppia)	Isomerase activity; unfolded protein binding (Protein folding)	NS	-2.43	-2.91*

Genes with the most consistent expression differences across both stages were found by comparing the gene lists of CS17 bat limbs vs. mouse with those created by a comparison with the CS16 bat limbs and the mouse (Fig. 4.19). These genes would be those that are not false positives as they are differentially expressed in two separate experiments. They would also be the genes that have prolonged differential expression in the bat limb type and therefore have a relevant function during development. Thirteen DE genes that were common to both bat embryonic stages were found (Fig. 4.9, blue in Table 4.8).

Twenty-one additional genes that were significantly differentially expressed between the CS17 HP and the mouse HP were also found (Fig. 4.9; Table 4.9). The majority of the differentially expressed genes had no sequence information for the bat and the similarity between the bat gene sequence and the Operon probe could not, therefore, be calculated. For the genes that did have bat sequences that were available, it was found that most showed a high similarity (>80%), with only two genes (*Syt13*, *Cdk3*) having lower similarities. No significant matches were made between the sequence for these probes and any other available bat sequences.

ii Intra-species comparison (Bat HP vs. Bat FP)

Six genes were found to be significantly differentially expressed between the bat CS16 HP and FP, with four of these having a greater signal in the bat HP (Fig. 4.9; Table 4.9).

Table 4.9: Summary of genes differentially expressed at a level 1.5 fold or higher between the CS16 bat hand and foot plates. With genes in bold being similarly regulated in the CS17 bat embryo. The OPERON ID (mouse gene accession number) is given for each probe, as well as its gene name (abbreviated name). † indicate when the gene associated with the probe was annotated through blasting. The gene ontology (GO) of the molecular and biological (in brackets) function is also shown. The fold change (FC) between the CS16 HP and FP are given. * signifies whether this change is statistically significant ($q < 0.01$).

OPERON ID (Accession No.)	Gene Name	GO: Molecular (Biological)	FC
M400017713 (A830011L22)	Meis homeobox 2 (Meis2)†	Transcriptor factor activity (Regulation of transcription, DNA-dependent	1.93*
M200002041 (NM_008273)	Homeo box D11 (Hoxd11)	Transcriptor factor activity (Skeletal development; pattern specification process)	1.79*
M400000669 (AC129016)	Nucleophosmin 1(Npm1)†	DNA binding (Protein localization; signal transduction by p53 class mediator)	1.59*
M400002268 (XM_485506)	adherens junction associated protein 1 (Ajap1)†	Protein binding (Cell adhesion)	1.56*
M400013559 (4933436D11)	Protein kinase C, delta (Prkcd)†	Protein kinase activity (Protein amino acid phosphorylation)	-2.03*
M300016981 (NM_172205.1)	Suprabasin (Sbsn)		-2.09*

In the bat CS17 HP and FP comparison fifteen genes were found to be differentially expressed with the majority of these (thirteen) having a greater expression in the bat HP (Table 4.10). The top differentially expressed candidate in all of the comparisons was the same. The gene *Meis2* shows the highest fold expression difference (FC = 9.91) in the CS17 HP of the bat compared to the mouse hand plate. It also has the highest fold expression difference (FC = 4.53) in the CS17 HP as compared to the FP and it has a reasonably high fold expression difference (FC = 1.93) in the CS16 HP as compared to the FP. This indicates that this gene has a prolonged expression in the developing bat autopods and that this expression is consistently higher in the bat HP as opposed to the FP. It can also be seen that this expression is higher in bat autopods in general than in the mouse HP. The gene *Hoxd11* is also found to be commonly differentially expressed in between the HP and the FP of both bat developmental stages.

To examine the relationship between differential expression of the CS17 and that of CS16 bat limbs, the top 5% (260 genes) of the differentially expressed gene lists (ranked by DEDS score) were compared. Highly ranked genes that are common to both stages would be of functional importance in the development of the respective autopod. *Meis2* and *Hoxd11* were the only genes in common that were found to be significantly differentially expressed, however five other genes were found to be highly ranked in both stages (Table 4.11). These included the important limb developmental genes *Hoxd10* and *Hoxd12*. *Hoxc6* was the only common gene found to have a greater expression in the bat FP as opposed to the bat HP.

Table 4.10: Summary of genes differentially expressed at a level 1.5 fold or higher between the CS17 bat hand and foot plates. With genes in bold being similarly regulated in the CS17 bat embryo The OPERON ID (mouse gene accession number) is given for each probe, as well as its gene name (abbreviated name). † are given when the gene associated with the probe was annotated through blasting. The gene ontology (GO) of the molecular and biological (in brackets) function is also shown. The fold change (FC) between the CS16 HP and FP are given. * signifies whether this change is statistically significant ($q < 0.01$).

OPERON ID (Accession No.)	Gene Name	GO: Molecular (Biological)	FC
M400017713 (A830011L22)	Meis homeobox 2 (Meis2)†	Transcriptor factor activity (Regulation of transcription)	4.53*
M200002041 (NM_008273)	Homeo box D11 (Hoxd11)	Transcriptor factor activity (Skeletal development; pattern specification process)	2.33*
M300005520 (NM_030725)	synaptotagmin XIII (Syt13)	Transporter activity (Vesicle-mediated transport)	2.03*
M200000546 (NM_010769)	Matrilin 1, cartilage matrix protein (Matn1)	Collagen binding (Matrix organization and biogenesis)	1.83*
M400018107 (AC117574)	RIKEN cDNA 4930428O21 gene		1.80*
M400005885 (AC121950)	Cancer susceptibility candidate 1 (Casc1)†		1.79*
M400014003 (BC048660)	RIKEN cDNA 1700125D06 gene		1.78*
M200015556 (NM_027165)	Cyclin-dependent kinase 3 (Cdk3)†	Nucleotide binding (Cell cycle, Cell division)	1.77*
M400002504 (XM_127674)	RIKEN cDNA 3110001K24 gene	Hydrolase activity	1.75*
M200003355 (NM_018755)	Plasma glutamate carboxypeptidase (Pgcp)	Peptidase activity (Proteolysis)	1.72*
M300009692 (XM_485063)	RIKEN cDNA 3110001I20 gene		1.69*
M400013993 (AC139295)	RIKEN cDNA 1700047M11 gene		1.68*
M300002857 (NM_013500)	Hyaluronan and proteoglycan link protein 1 (Hapln1)	Hyaluronic acid binding (Cell adhesion)	1.64*
M400007409 (BC093519)	RNA binding motif protein 10 (Rbm10)†	RNA binding	-1.76*
M400006624 (AC162898)	predicted gene, ENSMUSG00000059064†	Structural constituent of ribosome (Translation)	-1.97*

Table 4.11: Summary of common differentially expressed genes between the bat hand and foot plate of stages CS17 and CS16. The OPERON ID (mouse gene accession number) is given for each probe, as well as its gene name (abbreviated name). † are given when the gene associated with the probe was annotated through blasting. The gene ontology (GO) of the molecular and biological (in brackets) function is also shown. The fold change (FC) between the both the CS17 and CS16 HP and FP are given. * signifies whether this change is statistically significant ($q < 0.01$).

OPERON ID (Accession No.)	Gene Name	GO: Molecular (Biological)	FC 17	FC 16
M400017713 (A830011L22)	Meis homeobox 2 (Meis2)†	Transcriptor factor activity(Regulation of transcription)	4.53*	1.93*
M200002041 (NM_008273)	Homeo box D11 (Hoxd11)	Transcriptor factor activity (Skeletal development; pattern specification process)	2.33*	1.79*
M200007845 (NM_026821)	Uncharacterized protein C9orf150 homolog (D4Bwg0951e)		1.51	1.38
M200014937 (NM_008274)	Homeo box D12 (Hoxd12)	Transcriptor factor activity (Skeletal development; pattern specification process)	1.47	1.49
M200004888 (NM_013554)	Homeo box D10 (Hoxd10)	Transcriptor factor activity (Skeletal development; pattern specification process)	1.39	1.37
M400011601 (NM_027326)	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 3 (Mllt3)	(regulation of transcription; Anterior/posterior pattern formation)	1.37	1.35
M300000219 (NM_010465)	Homeo box C6 (Hoxc6)	Transcription factor activity (embryonic skeletal development; anterior/posterior pattern formation)	-1.08	-1.45

The gene lists from the DEDS analyses were ranked according to their DEDS score and scanned using Fatiscan to determine whether the functional annotations (GO molecular and biological processes) were asymmetrically distributed (Al-Shahrour et al., 2006). This second analysis step mines the data for functional similarities within blocks of each gene list when it is ordered by differential expression. There were many functional groups which were significantly over- or underrepresented in the comparison between the bat CS17 HP and the mouse E13.5 HP (Table 4.12). The most prevalent of these was the overrepresentation of GO annotations related to receptor activity, cell signalling and developmental processes in the bat hand plate and the underrepresentation of annotations related to proteolysis (adjusted $p < 0.05$). The same analysis done between the bat foot plate and the mouse hand plate revealed no significantly over- or underrepresented GO annotations for this comparison. This implies that the gene annotation information revealed in the first analysis reveals molecular or

biological processes that play a role in determining the bat hand plate morphology rather than being a result of species-specific differences.

Table 4.12: Gene Ontology (GO) annotations that are under- or overrepresented in the gene list of differentially expressed genes of the CS17 HP as compared to the mouse E13.5 HP. GO annotations in boldface are molecular processes, the others are biological processes.

GO Level	Overrepresented	Adj. p value	Underrepresented	Adj. p value
Level 3	**receptor activity (GO:0004872)**	**0.000**	**structural constituent of ribosome (GO:0003735)**	**0.000**
	extracellular matrix structural constituent (GO:0005201)	**0.011**	**hydrolase activity (GO:0016787)**	**0.008**
	cell communication (GO:0007154)	0.000	cellular metabolic process (GO:0044237)	0.000
	neurological process (GO:0050877)	0.000	primary metabolic process (GO:0044238)	0.000
	digestion (GO:0007586)	0.022	macromolecule metabolic process (GO:0043170)	0.000
			biosynthetic process (GO:0009058)	0.000
			catabolic process (GO:0009056)	0.010
			protein localization (GO:0008104)	0.019
Level 4	**transmembrane receptor activity (GO:0004888)**	**0.000**	protein metabolic process (GO:0019538)	0.000
	sensory perception (GO:0007600)	0.000	cellular macromolecule metabolic process (GO:0044260)	0.000
	signal transduction (GO:0007165)	0.000	cellular biosynthetic process (GO:0044249)	0.000
	anatomical structure morphogenesis (GO:0009653)	0.018	ribonucleoprotein complex biogenesis and assembly (GO:0022613)	0.003
	regulation of developmental process (GO:0050793)	0.025	macromolecule catabolic process (GO:0009057)	0.008
			cellular catabolic process (GO:0044248)	0.033
			biopolymer metabolic process (GO:0043283)	0.038
			regulation of binding (GO:0051098)	0.044
Level 5	**G-protein coupled receptor activity (GO:0004930)**	**0.000**	**transition metal ion binding (GO:0046914)**	**0.009**
	rRNA binding (GO:0019843)	**0.010**	macromolecule biosynthetic process (GO:0009059)	0.000
	transcription factor activity (GO:0003700)	**0.012**	cellular protein metabolic process (GO:0044267)	0.000
	sensory perception of chemical stimulus (GO:0007606)	0.000	ribosome biogenesis and assembly (GO:0042254)	0.000
	cell surface receptor linked signal transduction (GO:0007166)	0.000	DNA metabolic process (GO:0006259)	0.005
			cellular macromolecule catabolic process (GO:0044265)	0.020

Functional analyses between the bat hand and foot plates at stages CS17 and CS16 revealed fewer functional annotations that were asymmetrically distributed. In the CS17 HP endoribonuclease activity is overrepresented while nucleotide binding is underrepresented (Table 4.13). In the CS16 HP the structural constituents of ribosomes were overrepresented while ligase activity was underrepresented (Table 4.14). The isolation of significantly asymmetrically distributed functional classes indicates that there is additional information inherent in the dataset that could be obtained through appropriate data mining techniques. No functional pathways were found to be significantly different among all comparisons (adj. $p < 0.05$).

Table 4.13: Gene Ontology (GO) annotations that are under or overrepresented in the gene list of differentially expressed genes of the CS17 HP as compared to the CS17 FP. GO annotations in **boldface** are molecular processes, the others are biological processes

GO Level	Over represented	Adj. p value	Under represented	Adj. p value
Level 4			**nucleotide binding (GO:0000166)**	**0.013**
Level 7	**endonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 3'-phosphomonoesters (GO:0016894)**	**0.010**		
	endoribonuclease activity (GO:0004521)	**0.012**		
Level 9	**endoribonuclease activity, producing 3'-phosphomonoesters (GO:0016892)**	**0.003**		
Level 10	**pancreatic ribonuclease activity (GO:0004522)**	**0.011**		

Table 4.14: Gene Ontology (GO) annotations that are under- or overrepresented in the gene list of differentially expressed genes of the CS16 HP as compared to the CS16 FP. GO annotations in **boldface** are molecular processes, the others are biological processes.

GO Level	Overrepresented	Adj. p value	Underrepresented	Adj. p value
Level 3	**structural constituent of ribosome (GO:0003735)**	**0.028**		
Level 4			**ligase activity, forming carbon-oxygen bonds (GO:0016875)**	**0.034**
Level 5	DNA metabolic process (GO:0006259)	0.044		
Level 7			amino acid activation (GO:0043038)	0.016
Level 8			tRNA aminoacylation (GO:0043039)	0.019
			striated muscle cell differentiation (GO:0051146)	0.034
Level 9			tRNA aminoacylation for protein translation (GO:0006418)	0.010

4.3.9 Gene expression of limb developmental pathways

Additional data mining was done by manually selecting and visualising the array signal from genes known to be involved in limb development of both the mouse and the bat. These genes and their families were examined to determine if these changes were evident in the expression data and to mine information from associated genes. Gene expression data was extracted as the average single channel spot signal across all four arrays for each gene and is hereafter referred to as the signal. Retinoic acid (RA) has been show to play a pivotal role in the developing limb, playing an important role in the activation of genes involved in limb bud outgrowth and patterning (Mercader et al., 2000). Several of the genes involved in the synthesis, degradation and signalling of RA were examined (Fig. 4.20A).

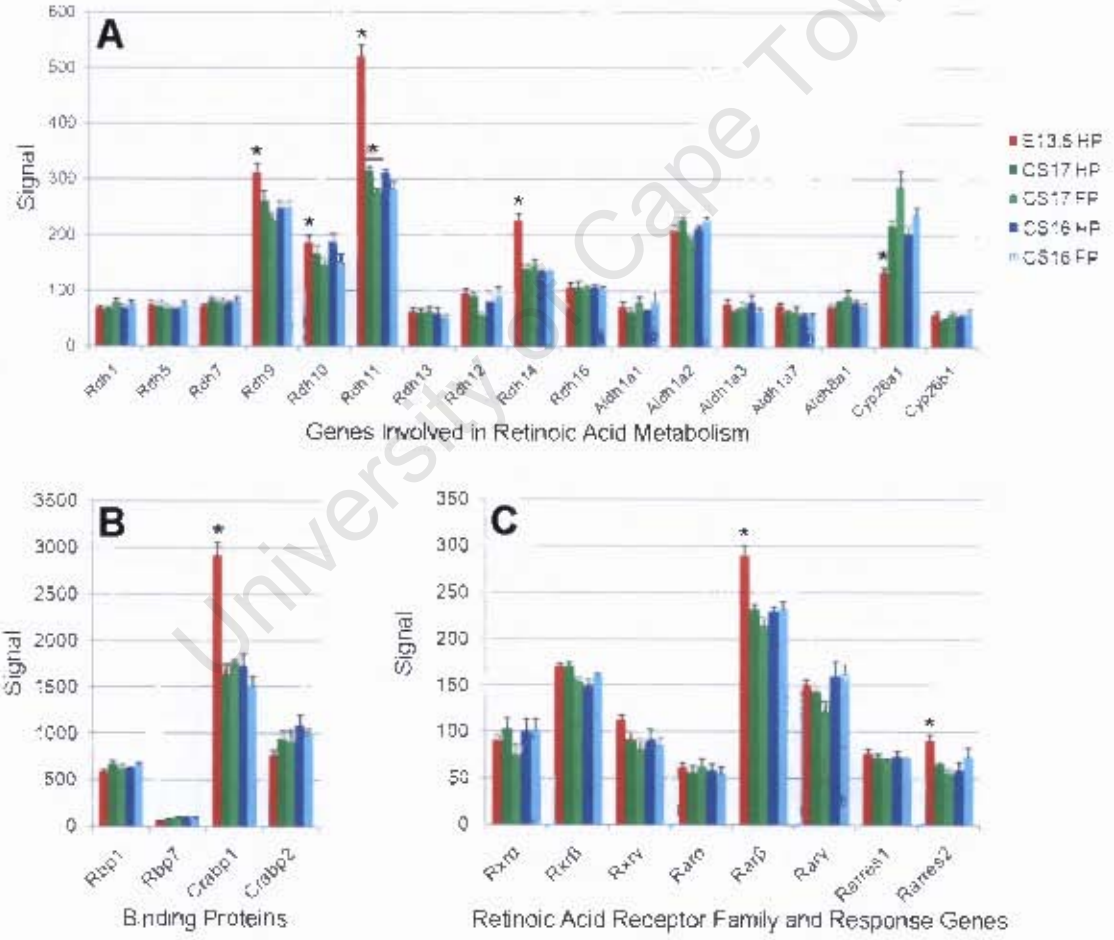


Fig. 4.20: Expression of genes involved in retinoic acid (RA) (A) metabolism, (B) binding and (C) signalling. Greater signal was found in the mouse HP as compared to the bat HP for many of these genes. Average signal from the mouse E13.5 HP transcripts (red) are compared to that of the CS17 bat HP (dark green). * above the mouse bar denotes that these two signals were significantly different (Mann-Whitney U, $p < 0.05$). The bat CS17 HP was also compared to the CS17 FP (light green) and the bat CS16 HP (dark blue) was compared to the CS16 FP (light blue). A bar across the two limb conditions with * above it denotes that these two signals were significantly different. Error bars represent the standard error (SE) across all the arrays in the analysis.

Some of the genes whose products are involved in the synthesis of RA were found to have a greater signal in the mouse HP (Fig. 4.20A), these included *Retinol dehydrogenase 9* (*Rdh9*; FC = 1.2), *Retinol dehydrogenase 10* (*Rdh10*; FC = 1.1), *Retinol dehydrogenase 11* (*Rdh11*; FC = 1.6) and *Retinol dehydrogenase 14* (*Rdh14*; FC = 1.6) (Fig. 4.20). *Rdh11* was also found to be significantly higher between the bat CS17 HP and FP (FC = 1.1). *Cytochrome P450, family 26, subfamily a, polypeptide 1* (*Cyp26a1*), a gene whose product is involved in breaking down RA into an inactive form (McCaffery and Simons, 2007), had a higher signal in the bat CS17 HP compared to the mouse HP (FC = 1.6). This would appear to indicate that there would be a higher synthesis of RA in the mouse HP as compared to the bat HP. Some of the genes involved in RA signalling were also found to have a higher signal in the mouse HP than in the bat HP (Fig. 4.20B & C). These included the RA binding protein, *Cellular retinoic acid binding protein I* (*Crabp1*; FC = 1.1), the RA receptor, *Retinoic acid receptor, beta* (*Rarβ*; FC = 1.2), and the retinoic acid receptor responder, *Retinoic acid receptor responder (tazarotene induced) 2* (*Rarres2*; FC = 1.4). No significant differences in signal were seen for these genes in either of the bat comparisons.

Previous research on bat limb development has found changes in the expression levels and patterns of several limb development genes. These include *Homeobox D13* (*Hoxd13*) (Cretekos et al., 2005), *Bone morphogenetic protein 2* (*Bmp2*) (Sears et al., 2006), *Fibroblast growth factor 8* (*Fgf8*) (Weatherbee et al., 2006) and *Paired related homeobox 1* (*Prrx1*) (Cretekos et al., 2008).

The *Hox A, C* and *D* clusters were examined. Significant differences were found between the mouse E13.5 HP and the bat CS17 HP (Mann-Whitney U test: $p < 0.05$). *Homeobox A9* (*Hoxa9*; FC = 1.9) and *Homeobox A11* (*Hoxa11*; FC = 4.2) showed a lower signal in the bat hand plate while *Homeobox A10* (*Hoxa10*; FC = 1.4) and *Homeobox A13* (*Hoxa13*; FC = 1.4) showed higher signal (Fig. 4.21A). There was significantly higher signal of *Hoxa11* in the CS16 FP as opposed to the HP (FC = 1.2). The signal from genes in the *Hox C* cluster was much lower than that of *A* or *D* (Fig. 4.21B). Though there were significant differences these were relatively low level with the only interesting pattern seen in *Homeobox C6* (*Hoxc6*; FC = 1.6). The signal from the mouse E13.5 HP was lower than that of the bat; however signal from the FP of both bat stages was higher than that of the HP (CS17 FC = 1.5; CS16 FC = 1.6).

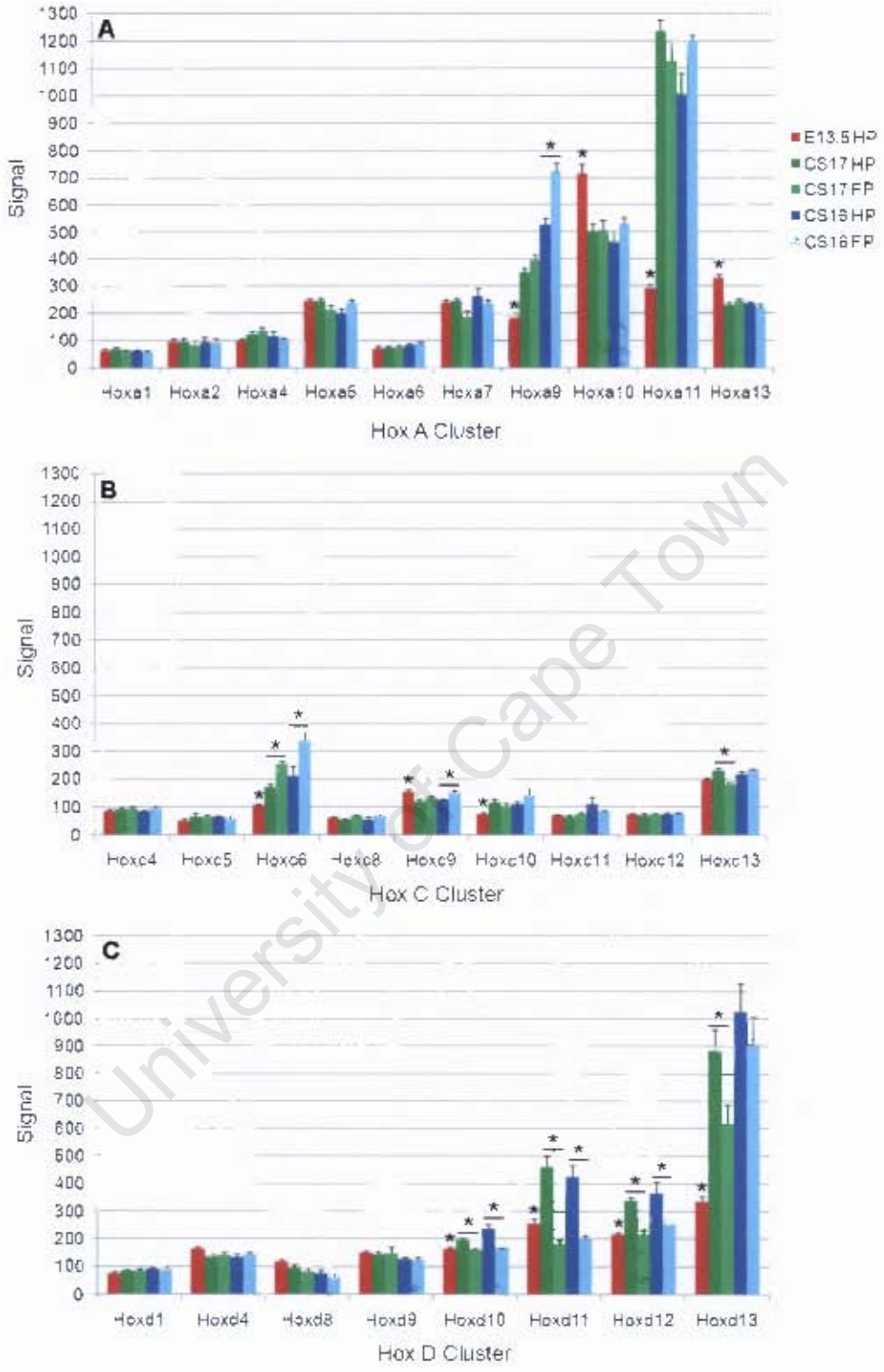


Fig. 4.21: Expression of genes belonging to the (A) *Hox A*, (B) *Hox C* and (C) *Hox D* clusters. Significant differences were found between the mouse and the bat for many of $Hoxa^{9-13}$ and $Hoxd^{9-13}$ genes, which are important in mediating limb developmental processes. Average signal from the mouse E13.5 HP transcripts (red) are compared to that of the CS17 bat HP (dark green). * above the mouse bar denotes that these two signals were significantly different ($p < 0.05$) The bat CS17 HP was also compared to the CS17 FP (light green) and the bat CS16 HP (dark blue) was compared to the CS16 FP (light blue). A bar across the two limb conditions with * above it denotes that these two signals were significantly different. Error bars represent the standard error (SE) across all the arrays in the analysis.

The *Hox D* cluster shows the clearest and most consistent pattern of differential expression, with *Homeobox D10, 11, 12 and 13* (*Hoxd10*, *11*, *12* and *13*) having a lower signal in the mouse E13.5 HP (*Hoxd10*: FC = 1.2, *Hoxd11*: FC = 1.8, *Hoxd12*: FC = 1.5, *Hoxd13*: FC = 2.6; Fig. 4.21C). The bat HP of both stages is consistently higher than the FP for the genes *Hoxd10*, *11* and *12*. (*Hoxd10*: CS17 FC = 1.2, CS16 FC = 1.4; *Hoxd11*: CS17 FC = 2.5, CS16 FC = 2.1; *Hox12*: CS17 FC = 1.6, CS16 FC = 1.4). For these genes, the signal in the mouse HP appears to be equivalent to that in the bat FP of both stages. The expression in the bat HPs was also notable for its similarity across both stages of development. *Hoxd13* shows significantly higher signal from the CS17 HP of the bat than from the FP (FC = 1.4), though this same pattern can be seen in CS16 there is high variation among the biological repeats and the differences were not found to be significant. Within this group *Hoxd10* has the lowest expression in the bat HP and *Hoxd13* the highest. However, the expression of *Hoxd12* is lower than that of *Hoxd11*.

Expression of members from the BMP signalling family was significantly different between the mouse HP and the bat CS17 HP (Fig. 4.22A).

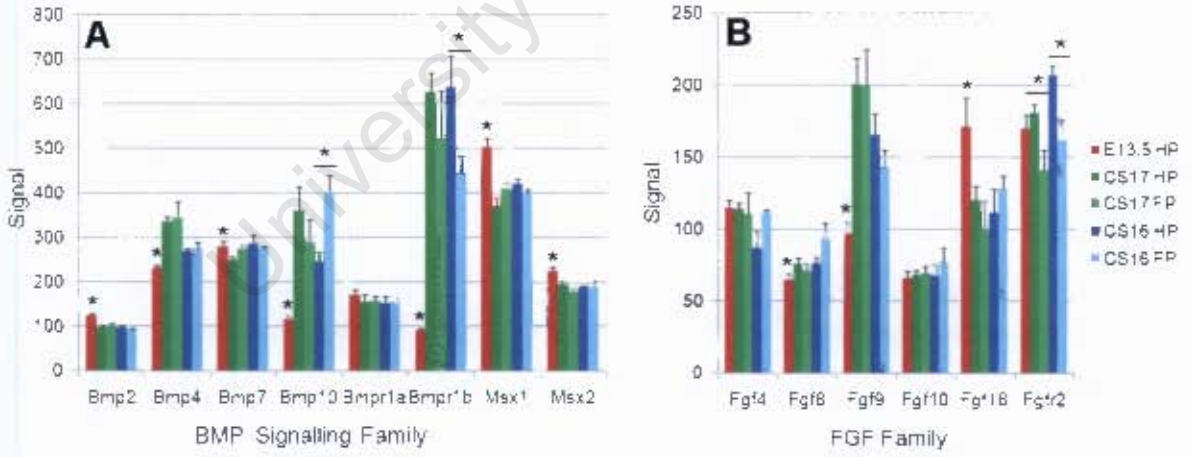


Fig. 4.22: Expression of (A) Bone morphogenic protein (*Bmp*) and the (B) Fibroblast growth factor (*Fgf*) signalling genes. (A) The *Bmp* genes showed relatively few differences in signal between the limb conditions. the most prominent differences occurred in *Bmp10* and *Bmpr1b* where the signal in the bat CS17 HP was much higher than that of the mouse. (B) Only *Fgf9* showed strong signal differences among the *Fgf* family, with a much lower signal form the mouse HP than form the bat CS17 HP. Average signal from the mouse E13.5 HP transcripts (red) are compared to that of the CS17 bat HP (dark green). * above the mouse bar denotes that these two signals were significantly different ($p < 0.05$). The bat CS17 HP was also compared to the CS17 FP (light green) and the bat CS16 HP (dark blue) was compared to the CS16 FP (light green). A bar across the two limb conditions with * above it denotes that these two signals were significantly different. Error bars represent the standard error (SE) across all the arrays in the analysis.

Bmp2 (FC = 1.1) and *Bmp7* (FC = 1.1) had a slightly higher signal in the mouse while *Bmp 4* (FC = 1.4) and *Bmp10* (FC = 3.1) and the receptor, *Bone morphogenetic protein receptor, type 1B* (*Bmpr1b*; FC = 6.8) was higher in the bat HP (Mann-Whitney U test, $p < 0.05$). The downstream effectors of BMP signalling *Homeobox msh-like 1* and *2* (*Msx1* FC = 1.4 and *Msx2* FC = 1.2) showed a higher expression in the mouse HP. Only two genes showed a significant difference between the bat limb conditions, *Bmp10* had a greater signal in the CS16 FP (FC = 1.6) while *Bmpr1b* had a greater signal in the CS16 HP (FC = 1.4).

The FGF family showed a surprisingly low signal among all the limb conditions (Fig. 4.22B). The difference in *Fgf8* signal between the bat and mouse was significant if minimal (FC = 1.2). The greatest differences between the mouse and the bat were seen in the low expression of *Fgf9* (FC = 2.1) in the mouse hand plate and the high expression of *Fgf18* (FC = 1.4) in the mouse hand plate. The only significant difference exhibited between the two bat limb types was found with the *Fgfr2* signal, with the HP having a greater signal than the FP (CS17 FC = 1.3; CS16 FC = 1.3) (Mann-Whitney U test, $p < 0.05$).

The signal from *Paired related homeobox 1* (*Prrx1*) and its associated genes was also examined (Fig. 4.23). *PTK protein tyrosine kinase 2* (*Ptk2*), an upstream activator of *Prrx1* expression (McKean et al., 2003) was found to have low expression with no significant differences between limb types. This was in contrast to the signal from *Prrx1* itself which was found to be significantly lower in the mouse HP (FC = 1.8) and the bat CS17 FP (FC = 1.4) than in the bat CS17 HP. No significant differences were found between the CS16 HP and FP signal even though higher signal was found in the HP and lower signal in the FP (FC = 1.1). *Paired related homeobox 2* (*Prrx2*) signal did not follow that of *Prrx1*, and no significant differences were found in the three comparisons. The signal from *Tenascin C* (*Tnc*), a gene activated by the expression of *Prrx1* (Jones et al., 2001), followed the same pattern as that of *Prrx1* with higher expression occurring in the bat CS17 HP than in either the mouse HP (FC = 1.6) or the bat CS17 FP (FC = 1.4) (Mann-Whitney U test, $p < 0.05$).

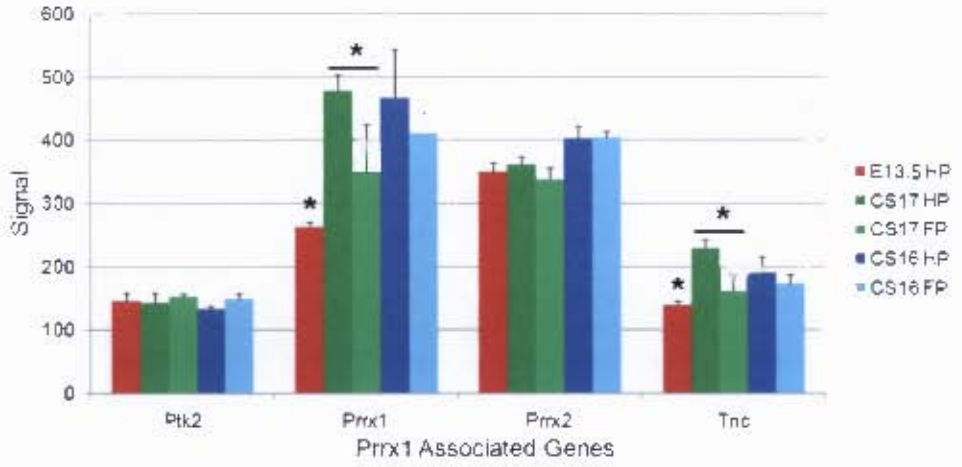


Fig. 4.23: Expression analysis of *Prrx1* associated genes showing that the bat CS17 HP has a significantly higher signal than that of the FP or the mouse HP from both *Prrx1* and its downstream target *Tnc*. Average signal from the mouse E13.5 HP transcripts (red) are compared to that of the bat (dark green. * above the mouse bar denotes that these two signals were significantly different ($p < 0.05$) The bat CS17 HP was also compared to the CS17 FP (light green) and the bat CS16 HP was compared to the CS16 FP. A bar across the two limb conditions with an asterisk above it denotes that these two signals were significantly different. Error bars represent the standard error (SE) across all the arrays in the analysis.

4.3.10 *Meis2* identified probe reflect novel transcript

The Operon probe M400017713 was shown to be significantly differentially expressed between all comparisons and was identified as corresponding to the *Meis2* gene, however a second probe (M400000987) was also identified as corresponding to the *Meis2* (*Mrg1*) gene. Blasting analysis confirmed that these two probes were both mapped to different positions on the *Meis2* gene. The first probe (M400017713) was 100% similar (without gaps) to the Riken clone: A830011L22, identified as *Meis2* from an E10.0 mouse neonate cortex full-length enriched cDNA library (Score = 130, E = $6e^{-28}$). It had the same similarity match for a mouse cloned genomic DNA sequence (clone RP23-162G11 on chromosome 2) that contained the 5′ end of the *Meis2* gene sequence. The second probe (M400000987) was 100% similar (without gaps) to mouse cloned genomic DNA sequence (clone: RP23-297K24 on chromosome 2) that contained the 3′ end of the *Meis2* gene sequence (Score = 99.6, E = $8e^{-19}$). Therefore the first probe signal (M400017713) was mapped to the 5′portion and the second probe (M400000987) to the 3′ portion of this gene (hereafter the probes will be referred to as 5′-*Meis2* and 3′-*Meis2* respectively). The signals of these two probes should both bind specifically to the *Meis2* gene, however they showed dramatically different signal, with the first probe (M400017713) having signal that was more than eight times that of the second probe (M400000987) in the CS17 HP and nearly two times higher in the other bat limb conditions (Fig. 4.24). These two probes had similar expression levels in the mouse HP.

Meis1 and *Meis3* showed no significant differences in signal between the different species and limb conditions (Fig. 4.24). The 3'-*Meis2* probe signal was slightly, if significantly, higher in the CS17 HP than either the mouse E13.5 HP (FC = 1.7) or the CS17 FP (FC = 1.6). A gene, *Pbx1*, with expression that is associated with the *Meis* gene family showed a similar signal level to that of the 3'-*Meis2* probe, with greater signal occurring in the bat CS17 HP than in either the mouse E13.5 HP (FC = 2.5)or the bat CS17 FP (FC = 1.4) (Fig. 4.24). The signal from the 5'-*Meis2* probe had significantly higher expression in the CS17 bat hand than in the mouse E13.5 HP (FC = 13.1) and in the CS17 FP (FC = 7.0). This gene was also highly expressed in the CS16 HP than in the CS16 FP (FC = 2.2).

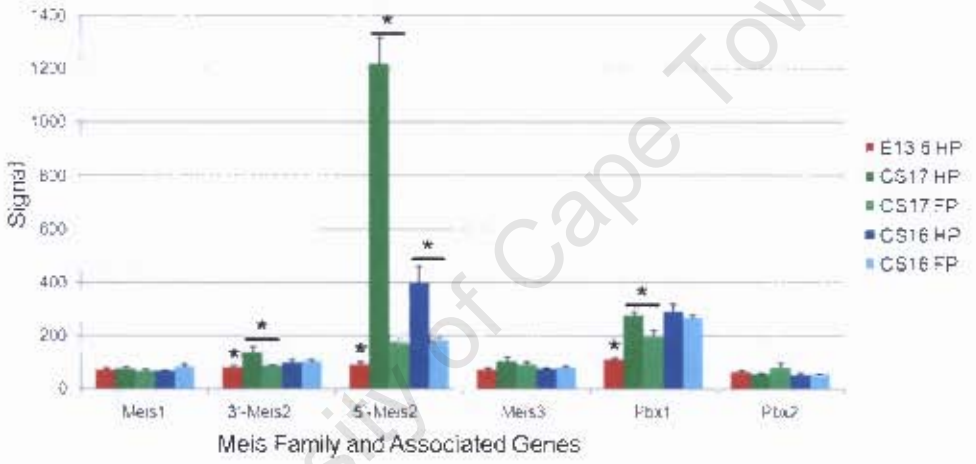


Fig. 4.24: Expression of the *Meis* gene family and their associated genes showing the significantly different signal of 3'-*Meis2* from the annotated gene 5'-*Meis2*. Average signal from the mouse E13.5 HP transcripts (red) are compared to that of the CS17 bat HP (dark green). * above the mouse bar denotes that these two signals were significantly different ($p < 0.05$). The bat CS17 HP was also compared to the CS17 FP (light green) and the bat CS16 HP (dark blue) was compared to the CS16 FP (light blue). A bar across the two limb conditions with * above it denotes that these two signals were significantly different. Error bars represent the standard error (SE) across all the arrays in the analysis.

The 5'-*Meis2* probe had 23 high similarity (>93%) matches (Table 4.15). These were the predicted *Meis2* gene sequences of several species, including the cow (*Bos Taurus*) the horse (*Equus caballus*), the dog (*Canis familiaris*) and the chimp (*Pan troglodytes*) as well as the *Meis2* gene sequence of the human (*Homo sapiens*) (Score > 107, E < $3e^{-21}$). The Riken clone: A830011L22 was the only *Meis2* mouse sequence match for this probe. The 3'-*Meis2* probe had 79 close matches (100%) to *Meis2* related sequences (the top 23 of which are given in Table 4.15).

Table 4.15: Summary of the top 23 closest match sequences for the *Meis2* probes ranked according to their megablast score (not shown). Accession numbers and the common name of the species (Name) it was found in is given along with a description of the sequence match. * indicates if it is a predicted sequence and its percentage similarity (%) is given. Sequences highlighted in blue indicate those that have similarities to both the *5'-Meis2 and the 3'-Meis2* probe

5'-*Meis2* Probe (M400017713)				**3'-*Meis2* Probe (M400000987)**			
Accession	Name	Description	%	Accession	Name	Description	%
AK043601.1	Mouse	Riken cDNA clone	100	AK300247.1	Human	cDNA clone	100
AL772135.14	Mouse	DNA clone	100	AK294411.1	Human	cDNA clone	100
XM_001787410.1	Cow	*Meis2*-001 mRNA*	95	AB463111.1	Human	DNA clone	100
XM_001503626.2	Horse	*Meis2* mRNA*	95	XM_001787419.1	Cow	*Meis2*-002 mRNA*	100
XM_852181.1	Dog	*Meis2*-010 mRNA*	95	XM_001787410.1	Cow	*Meis2*-001 mRNA*	100
XM_852139.1	Dog	*Meis2*-009 mRNA*	95	XM_001503626.2	Horse	*Meis2*-mRNA*	100
XM_852092.1	Dog	*Meis2*-008 mRNA*	95	XM_001925885.1	Pig	*Meis2*-002 mRNA*	100
XM_843724.1	Dog	*Meis2*-001 mRNA*	95	EU831754.1	Human	Synthetic clone	100
XM_851841.1	Dog	*Meis2*-003 mRNA*	95	EU831831.1	Human	Synthetic clone	100
XM_851804.1	Dog	*Meis2*-002 mRNA*	95	EU831824.1	Human	Synthetic clone	100
XM_001136864.1	Chimp	*Meis2*-002 mRNA*	94	CU690675.1	Human	Synthetic clone	100
XM_001136939.1	Chimp	*Meis2*-003mRNA*	94	CU674745.1	Human	Synthetic clone	100
XM_001136789.1	Chimp	*Meis2*-001 mRNA*	94	EU446635.1	Human	Synthetic clone	100
XM_001137088.1	Chimp	*Meis2*-004 mRNA*	94	AK311546.1	Human	cDNA clone	100
XM_001137495.1	Chimp	*Meis2*-005 mRNA*	94	AB174332.1	Monkey	cDNA clone	100
XM_001137572.1	Chimp	*Meis2*-006 mRNA*	94	AK234783.1	Pig	cDNA clone	100
NM_170677.2	Human	*Meis2*-a mRNA	94	XM_001136939.1	Chimp	*Meis2*-003 mRNA*	100
NM_170676.2	Human	*Meis2*-d mRNA	94	XM_001136789.1	Chimp	*Meis2*-001 mRNA*	100
NM_170675.2	Human	*Meis2*-c mRNA	94	XM_001137088.1	Chimp	*Meis2*-004 mRNA*	100
NM_170674.2	Human	*Meis2*-b mRNA	94	XM_001137495.1	Chimp	*Meis2*-005 mRNA*	100
AK056216.1	Human	cDNA clone	94	XM_001137572.1	Chimp	*Meis2*-005 mRNA*	100
AC087283.10	Human	DNA clone	94	XM_510290.2	Chimp	*Meis2*-005 mRNA*	100
AC078909.7	Human	DNA clone	94	AK226072.1	Human	*Meis2*-c mRNA*	100

These included *Meis2* transcript sequences from the above species in addition to those of the pig (*Sus scrofa*), the Long-tailed Macaque (*Macaca fascicularis*) and the Rhesus monkey (*Macaca mulatta*) (Score = 99.6, E = $8e^{-19}$). Thirteen mouse transcript sequences had matches to this probe, but, the RIKEN cDNA identified in the *5'-Meis2* blasting analysis was not among them. Eighteen transcript sequences were common between the blast analyses for both probes. No bat transcripts were found in these blasts and no significant matches for either probe were found when blasting the sequence against the nucleotide data base for Chiroptera (taxid: 9397). The high similarity of both probe binding site sequences, among a

variety of species, indicates that it is likely that both probes would bind to the bat homologue of the *Meis2* gene transcript.

Closer examination of an alignment of *Meis2* sequences across several species, including the predicted bat mRNA sequences from *Pteropus vampyrus* and *Myotis lucifugus* (predicted from human transcript sequences, Ensembl Genome Browser, ver. 52), showed that the 3′-*Meis2* probe would specifically hybridise over a highly conserved sequence close to the 3′ end of the coding region of the cDNA molecule, within exon 14 (Fig. 4.25 & Fig. 4.26). The 5′-*Meis2* probe would hybridise over a sequence within the 5′-UTR of the cDNA molecule in exon 1. This region was not found in the cDNA sequences of many of the species examined, differences that were found between the sequences of the closely related human and mouse suggest that this non-coding region may be less conserved than that of the 3′-*Meis2* probe (Fig. 4.26).

The mouse *Meis2* gene has 16 exons and 9 transcript variants (Ensemble Genome Browser, ver. 52). The 3′-*Meis2* probe was designed to the conserved homeodomain-like domain (Fig. 4.25). It is a common oligo and should represent all the transcripts of this gene (OPERON, datasheet). However, based on the probe binding sites, the *Meis2* transcripts that would give signal on the array can be categorised into three different groups. Type 1 transcripts would hybridise to both probes, and their expression would be represented by the signal from both 5′-*Meis2* and 3′-*Meis2*. Only one Ensembl mouse transcript (*Meis2*-001) was found to belong to this category (Fig. 4.24 & Fig. 4.25). Type 2 transcripts would be bind only to the 3′-*Meis2* probe and their expression would be represented by the 3′-*Meis2* signal only. The 8 other Ensemble mouse transcripts (*Meis2*-002, -003, -006, -011, -012, -013, -014 and -201) were found to belong to this category (Fig. 4.24 & Fig. 4.25). Type 3 transcripts would bind selectively to the 5′-*Meis2* probe and would their expression would be represented by the 5′-*Meis2* signal only. No Ensembl mouse transcripts in were type 3 transcripts, however a RIKEN full-length cDNA clone was found to belong to this category (Fig. 4.24, Fig. 4.25 & Fig. 4.26). While the type 1 and type 2 transcripts are predicted to encode the full-length MEIS2 protein, the type 3 transcript is predicted to encode a truncated protein that lacks the homeodomain-like domain (Fig. 4.25). As the signal of interest occurred only for the 5′-*Meis2* probe, it is expected that it should reflect the abundance of a type 3 transcript, identified as the RIKEN clone A830011L22.

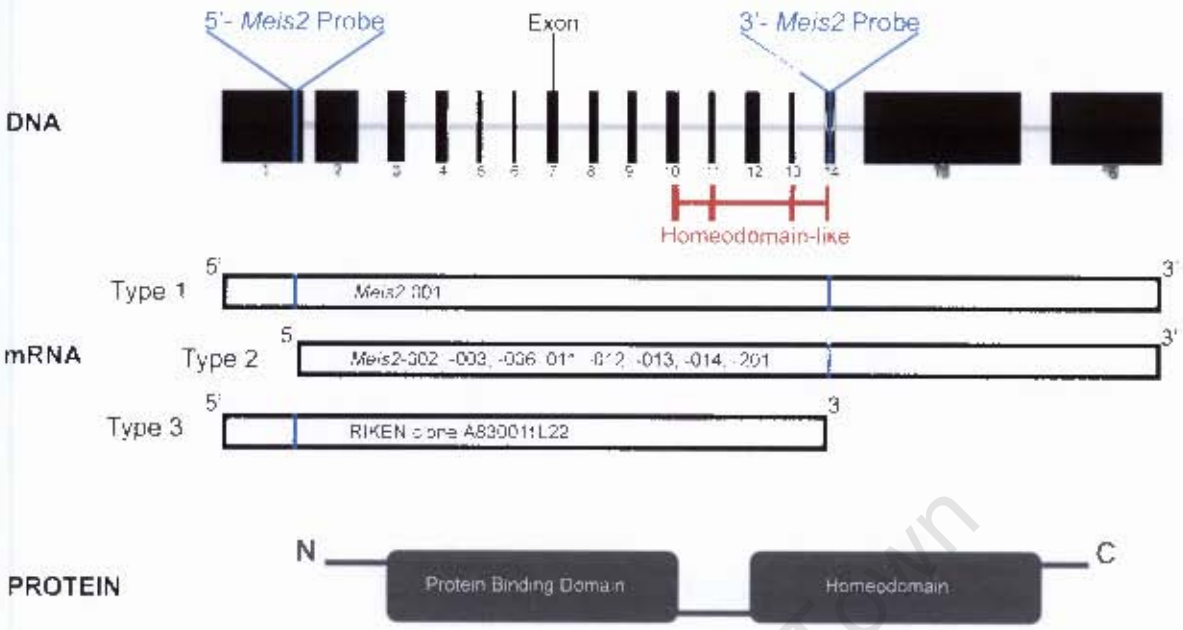


Fig. 4.25: The regions over which the 5'- and the 3'-Meis2 probes would bind are shown on a schematic of the gene. The three different transcript types that would give a signal on the microarray are shown, with the transcript numbers of their respective mouse sequences shown inset. These transcripts are known to code for a protein with a protein binding domain and a homeodomain.

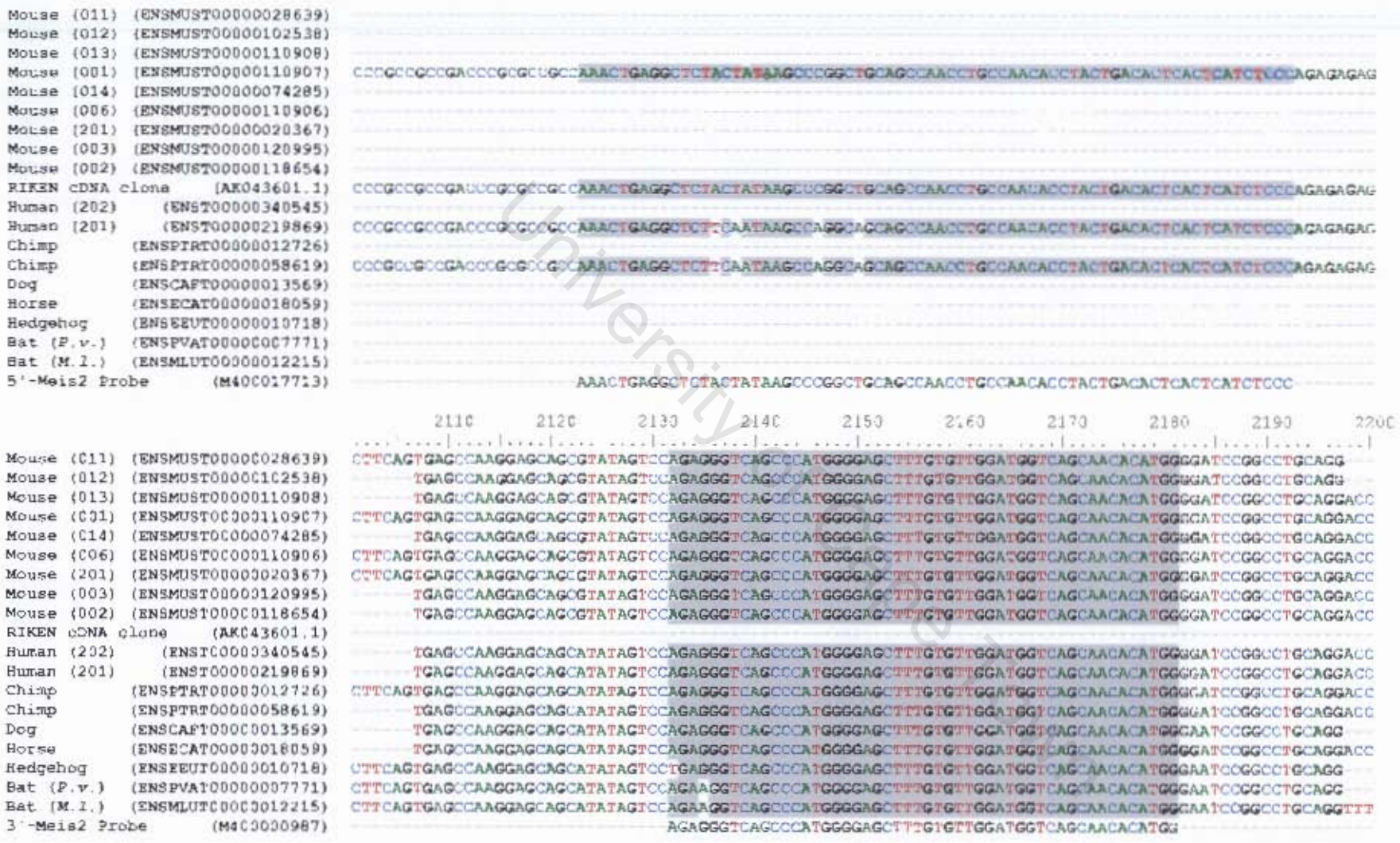


Fig. 4.26: The region over which the 5'- and the 3'-Meis2 probes would bind are shown on an alignment of actual transcript sequences of different species. The sequences of the probe binding sites show high similarity to both OPERON probes (greyed areas)

4.4 DISCUSSION

4.4.1 Differential expression analysis reveals novel candidate genes

The powerful, high throughput technique of microarray analysis has been used to gain additional insight into processes that occur during limb development in well characterised organisms such as the mouse (Shou et al., 2005). Although independent verification of the data remains to be done, this analysis technique proved to be equally useful when examining limb development in the bat, a non-model organism. Bat limb development is a recent field of research and there is a relative dearth of information on the species-specific processes that occur to make the bat wing a morphologically distinct structure. This study provides a fresh perspective on the field, identifying thirty-four robust, significantly different candidate genes. The majority of these genes were overexpressed in the bat hand plate as opposed to the mouse hand plate, and none had been previously characterised in bat limb developmental studies. Not only does this study point to the potential importance of these genes in shaping the bat wing, it also identifies them as having a role in limb development.

Three genes: *Meis homeobox 2* (*Meis2*), *Forkhead box K2* (*Foxk2*), and *Nuclear factor I/B* (*Nfib*), represented the most interesting and robust candidates as they are transcription factors that had a significantly higher expression in both the CS17 and CS16 hand plates as opposed to the E13.5 mouse hand plate. The annotated gene *Meis2* represented the most striking candidate gene among all analyses and is discussed further in Section 4.4.7. The *Foxk2* transcription factor is important for development and cell survival. It has previously been reported to be expressed in several epithelial tissues (in the kidneys, the intestines and the olfactory system) of the E18.5 mouse, often in regions of continued cellular proliferation (Wijchers, 2005). This gene has not been associated with limb developmental processes in the literature. However, it is expressed in both the fore- and hindlimb of the E10.5 mouse (Gray et al., 2004; Mouse Genome Informatics Web; Smith et al., 2007) . Its overexpression in the bat as opposed to the mouse hand plate (FC = 3.1) indicates that it may play a role in the continued proliferation of cells in the interdigital epithelial tissue of the bat hand plate. *Nuclear factor I/B* is a transcription factor that is expressed in the interdigital regions of the E14.5 mouse autopods (Chaudhry et al., 1997). It plays an important role in cell proliferation and/or apoptosis in lung and brain development (Steele-Perkins et al., 2005). There is a high expression of this gene in the bat as opposed to the mouse hand plate (FC = 2.9) indicating

4.4.2 Limb developmental pathways are conserved between limb types

The high level of correlation found between the hand and foot array signal of the bat at different stages of development suggests that very few differences in gene expression occur among these tissues. This supports the theory that there may be conservation in the gene expression levels of autopod developmental genes, with a very small number of genes controlling bat limb type specification (Sears, 2008). This is validated by gene expression studies on the mouse embryo were very few differences were found between the E11.0 (Margulies et al., 2001) and the E12.5 hand and foot plate (Shou et al., 2005). Relatively few genes have also been found to be differentially expressed between WT and mutant mouse embryonic limbs exhibiting either phocomelia (Qin et al., 2002) or a reduction in size (Cobb and Duboule, 2005). This indicates the limb developmental pathways are fairly robust to permutations, even those that cause dramatic morphological changes. Only fifteen genes were differentially expressed between the CS17 hand and foot plates and only six genes differentially expressed between the CS16 hand and foot plate. Differential expression of only a small subset of genes suggests that large scale genetic changes are driven by a small group of genes. These may control regulatory processes allowing their expression to be translated into downstream gene expression pattern changes or very low level gene expression differences between the fore- and hindlimb. This low incidence of differential expression within bat limbs implies that genes identified through stringent statistical testing should be strong candidate genes. Overall this finding indicated that the genetic differences between fore and hind limb development is subtle, with small numbers of differentially expressed genes between different limb types appearing to be an inherent property of gene expression in vertebrate limb development.

Within the differentially expressed genes there was a predominance of low fold changes, especially within the bat limb comparisons. This has been noted in several other limb microarray experiments (Cobb and Duboule, 2005; Qin et al., 2002; Shou et al., 2005) and can be attributed to the heterogeneous nature of the developing limb, with the presence of many different cell types in the autopod tissue. Therefore only modest fold changes are seen when the expression of genes within the whole limb or autopod is averaged in the microarray experiment (Qin et al., 2002).

4.4.3 The retinoic acid pathway provides a new perspective on bat limb development

Retinoic Acid (RA) is an important morphogen involved in limb development. It plays a role in initial A-P axis patterning, growth events and interdigital apoptosis (Morriss-Kay and Sokolova, 1996). Early stage RA mutants have truncated limbs and shortened long-bones (Niederreither et al., 2002). However an overexposure of RA in the developing limb can also create this phenotype (Kochhar, 1973). This indicates that RA levels require a fine balance and the regulation of this may be pivotal in controlling limb developmental processes. An intriguing phenotype of RA mutants is that, the reduction in limb length differs between the fore- and the hindlimb (Niederreither et al., 2002), indicating that there is some factor, different between their environments, which allows the hindlimb to be unaffected by the loss of RA. This raises the possibility that the modulation of RA activity may naturally result in forelimb specific morphological changes. During limb bud initiation and outgrowth a gradient of RA occurs from the proximal portion (highest) to the distal edge (lowest) of the limb bud. This gradient of RA, and gene products that it activates, specify the P-D axis at this stage.

Changes in RA concentration would be reflected by changes in the expression of genes that regulate its synthesis and degradation. The retinol dehydrogenases (RDH) serve to convert retinol to retinal while the aldehyde dehydrogenases (ALDH) oxidise retinal to RA. The retinol binding proteins (RBP) bind to retinol within the cell and mediate the efficient conversion of retinol to RA, and the cellular retinoic acid binding proteins (CRABP) bind retinoic acid. These binding proteins mediate the interactions of the ligand within the cell in some cases promoting its metabolism into an active form and in other case sequestering it removing the active form. Retinoic acid is also actively degraded by an enzyme, CYP26 (see Napouli, (1996) for a review). Therefore the abundance of active RA is regulated through a balance of the above gene products (McCaffery and Simons, 2007). The transcripts of these genes are present in the interdigital mesenchyme and RA in this area plays a role in generating and maintaining digital separation (Niederreither et al., 2002). Though there were no large differences between the signals of the bat hand and foot for the above genes, there were large differences between the mouse hand and the CS17 bat hand. Four genes involved in RA synthesis were found to have significantly higher expression in the mouse hand (*Rdh9* FC = 1.2; *Rdh10* FC = 1.1; *Rdh11* FC = 1.6; and *Rdh14* FC = 1.6). This together with the fact

that there was a lower signal from the RA degrading enzyme gene (*Cyp26a1* FC = 1.6) indicates that RA levels are higher in the mouse than in the bat hand plate. The RA binding protein *Crabp1* also had a slightly higher signal in the mouse hand (FC = 1.1). The concentration of this protein has been shown to affect the amount and type of RA that is in the cell (Boylan and Gudas, 1992). Its activity is ligand specific and therefore it is unknown what effect an increase in its abundance would have on the levels of RA. If active RA was less abundant in the bat hand plate than the mouse hand plate this could result in the retention of the interdigital tissue in the bat hand plate. However, the signal from the bat foot plate, which undergoes interdigital regression, is not significantly different to that of the bat hand plate. This indicates that either this difference in expression is a species-specific one that does not play a role in specifying the unique morphology of the bat hand plate or that other factors in addition to RA may be involved in the regression of the tissues in the foot plate. It has been shown that, in some cases, RA induced mutations are forelimb specific (Niederreither et al., 2002). If there is some RA redundancy in the hindlimb that can rescue these mutants then this would imply that the reduction of RA through more conventional pathways can be compensated for, resulting in a normal phenotype in the hindlimb. The possibility that low levels of RA may occur in the bat hand plate which results in interdigital retention is an interesting one and should be experimentally tested.

4.4.4 The 5′ *Hoxd* genes show co regulation in the bat hand plate

Hoxd gene activity is essential for limb developmental processes, the modulation of their expression has been shown to be important regulator in the evolution of diverse vertebrate limbs (Gonzalez et al., 2007; Ray and Capecchi, 2008; Reno et al., 2008). During limb bud formation this cluster of genes is temporally regulated in a collinear fashion with expression of the more 3′ genes (beginning with *Hoxd1*) preceding that of the more 5′ genes (ending with *Hoxd13*) (Tarchini and Duboule, 2006). A second wave expression of the 5′ *Hoxd* genes (*Hoxd10*, *11*, *12* and *13*) occurs in the mouse autopod. These genes are spatially regulated in a reverse collinear manner with a high expression of *Hoxd13* occurring in all digits and progressively lower expressions of the three other genes occurring in the posterior four digits (excluding digit one) (Montavon et al., 2008). This expression pattern has been shown to correspond to two modular developmental regions in the autopod, with digit one representing a single growth module, independent from that of the posterior four digits, in the forelimb of the hominoids (Reno et al., 2008). These modules allow the independent modification of digit lengths that cause the variation of digit proportions seen among the hominoids (Reno et

al., 2008). These also correspond to the regions of differential elongation in the bat wing, with the posterior four digits being relatively elongated while the thumb is similar in length to the digits of the foot (Ray and Capecchi, 2008). It is notable that these 5′ *Hoxd* genes were all significantly higher in the CS17 bat hand plate as opposed to both the foot plate and the mouse hand plate (*Hoxd10*: FC = 1.2, *Hoxd11*: FC = 1.8, *Hoxd12*: FC = 1.5, *Hoxd13*: FC = 2.6). *Hoxd13*, in particular, showed a much higher level of expression in the bat hand plate, a feature which could not be distinguished in previous gene expression pattern studies (Chen et al., 2005; Ray and Capecchi, 2008). In addition to this, *Hoxd10* to *12* had similar expression across both stages of bat development examined (*Hoxd10*: CS17 FC = 1.2, CS16 FC = 1.4; *Hoxd11*: CS17 FC = 2.5, CS16 FC = 2.1; *Hox12*: CS17 FC = 1.6, CS16 FC = 1.4), indicating a persistence of differential expression. It is interesting to note that the deletion of the entire *Hoxd* cluster leads to reduced digits with and unchanged autopod patterning (Zakany and Duboule, 1996; Zakany et al., 1997). However, the functional relevance of 5′*Hoxd* overexpression is difficult to clarify; with loss of function mutations exhibiting a loss or reduction of skeletal elements along the P-D axis and gain of function mutations showing patterning defects along the A-P axis (see Zakany and Duboule (2007) for a review). The role of these *Hox* genes in the formation of both of these limb axes is mediated by their interactions with *Sonic hedgehog* (*Shh*) and the Zone of Polarising Activity (ZPA) (Capellini et al., 2006; Tarchini et al., 2006). This gene has recently been found to have a novel second wave of expression in the bat hand and foot plate as compared to the mouse (Hockman et al., 2008),and, as *Hox* genes have been proposed to be upstream regulators of *Shh* (Zakany et al., 2004), this may be mediated by the overexpression of the 5′ *Hox* genes in the bat hand plates.

Only *Hoxd11* was significantly differentially expressed between both the bat hand and foot stages of CS17 and CS16. This gene has also been found to be differentially expressed between the hand plate and foot plate of the mouse (E12.5, FC = 1.37) indicating that it may play a role in specifying limb type identity (Shou et al., 2005). However the fold changes noted between the bat hand and foot (CS17 FC = 2.5, CS16 FC = 2.1) were much higher than that found in the mouse, with the average signal from the bat foot plate of both stages being similar to that of the mouse E13.5 hand plate. Duplications of this gene in the mouse cause a small, but significant increase in the length of the forelimb digits (Boulet and Capecchi, 2002). This gene shares a semi-redundant function with *Hoxa11*, another *Hox* gene that was found to be overexpressed in the CS17 bat hand plate as compared to that of the mouse (FC =

4.2). The deletion of both of these genes in the mouse leads to the delay of maturation of cartilage templates and the failure to establish growth plates in the limbs. As a result the skeletal elements are disorganised and shortened (Boulet and Capecchi, 2004). In single mutants the abnormal growth plate structure persists through to post-natal development, leading to shortened long bones (radius and ulna) (Davis et al., 1995). Overexpression of both of these genes in the bat hand plate may modify the growth plate structure leading to the elongation of the digit skeletal elements during development. These modifications in the growth plate structure would persist through to post-natal development, accounting for the continuous and rapid proliferation and differentiation events that have been found to elongate the bat forelimb skeletal elements (Farnum et al., 2008a; 2008b).

The 5′ *Hoxd* genes were the only gene group that were co-regulated in such a similar manner and the similarities in expression levels among these genes leads to the conclusion that an alteration in their co-ordinated transcriptional activation may be responsible for their common upregulation in the bat hand. Transcription of these genes in the limb is controlled by two enhancer sequences: (i) the global control region (GCR) which is conserved among all vertebrates (Spitz et al., 2003), and (ii) *Prox*, which is found between the GCR and the 5′ *Hoxd* genes (Gonzalez et al., 2007). These sequences have been shown to drive 5′ *Hoxd* gene expression in tetrapod limbs in a complementary fashion. An examination of the GCR of two bat species (*Rhinolophus ferrumequinum* and *Myotis lucifugus*) showed bat specific sequence differences, with several nucleotide differences occurring in the more conserved CsA and CsB regions (Ray and Capecchi, 2008). Further, transgenic work found that these sequence changes in the GCR were translated into novel enhancer activity domains in the mouse brain, the zeugopod and the stylopod (Ray and Capecchi, 2008). This study illustrates that sequence changes in the bat GCR can lead to novel expression domains of the *Hoxd* genes. Enhancer sequence changes may result in species-specific differences in transcript abundance of the 5′ *Hoxd* genes but it cannot explain the higher expression of these genes in the bat hand plate as compared to the foot plate. However, *Hoxd11* is expressed at significantly higher levels in the mouse hand plate (Shou et al., 2005) indicating that additional factors must regulate the expression of these genes between the fore- and hindlimb. The fact that *Hoxd11* is significantly differentially expressed within all the limb type comparisons and that the 5′ *Hoxd* genes ($Hoxd^{10\text{-}13}$) show a similar pattern of regulation, is a strong indication that these genes play important roles in mediating the molecular processes of bat limb development.

4.4.5 *Bmp* and *Fgf* candidate genes identified in bat limb development

The Bone morphogenic protein (BMP) and Fibroblast growth factor (FGF) families and their cofactors have both been the subject of expression studies in bats (Sears et al., 2006; Weatherbee et al., 2006), as both BMPs and FGFs play an important role in axis specification, patterning and apoptotic events in the developing mouse limb.

The expression of *Bmp*s and their downstream targets, *Msx1* and *2*, in the interdigital tissues of the mouse has been shown to be related to the regression of this tissue (Chen and Zhao, 1998). An examination of *Bmp2*, *4* and *7* expression in the developing bat wing at CS16 and CS17 showed that these genes were expressed both in the hand and in the foot plate, even though only latter experience interdigital regression (Weatherbee et al., 2006). The downstream targets of these genes, *Msx1* and *2* showed a similar pattern, with high expression occurring in the interdigits of all the bat autopods. In this microarray study only small differences were seen in the expression of these genes, with slightly higher signal in the mouse hand plate than that of the bat (*Bmp2*: FC = 1.1; *Bmp7*: FC = 1.1; *Msx1*: FC = 1.4; *Msx2*: FC = 1.2). *Bmp4* had slightly lower expression in the mouse hand plate that of the bat (FC = 1.4). The slightly higher expression of these genes indicates that apoptotic events should be more prevalent in the mouse hand plate than that of the bat. However, no differences were seen among any of the bat autopods. This corroborates previous gene expression work on the bat (Weatherbee et al., 2006). Sears et al. (2006), found that *Bmp2* (but not *Bmp4* or *7*) was overexpressed in CS20 bat metacarpals as opposed to those of the equivalent mouse stage. This microarray experiment indicates that any differences in *Bmp2* expression at CS20 can not be attributable to promoter based changes as suggested by Sears et al. (2006) as these should result in the overexpression of *Bmp2* in a stage independent manner.

Though interdigital regression does not occur in the bat hand plate, genes involved in apoptotic events are found to be expressed in these regions. These signals are not involved in the thinning of the interdigital mesenchyme as cell death assays have found little cell death in the interdigital regions of the bat hand plate (Weatherbee et al., 2006). This indicates that there is an additional mechanism that allows the maintenance of this tissue in the bat hand plate in spite of these apoptotic signals (Weatherbee et al., 2006). It has been proposed that a mechanism of reduced *Bmp* signalling in conjunction with increased *Fgf* signalling in the interdigital regions of the bat forelimb may be involved in the formation of the interdigital

webbing. A novel expression domain of *Fgf8* reported in the interdigital tissue of the developing bat hand plate at both CS17 and CS16 was proposed to promote cell survival in the presence of *Bmp* expression, therefore preventing apoptosis in this region (Weatherbee et al., 2006). No differences were found in the expression pattern of *Fgf8* among the bat autopods in the present study. This gene had an overall weak signal from all limb types and though it was significantly higher in the bat hand this was a very low-level difference. Though the expression domains of *Fgf* and *Bmp* may be correlated (Weatherbee et al., 2006) this is not supported by the relative signal of these genes in the bat autopod, with slightly higher expression of the *Bmp*s and lower expression of *Fgf*s occurring in the bat hand plate as compared to the mouse. The only member of the *Fgf* family examined to show a significantly different expression among the bat autopods was the FGF receptor, *Fgfr2*, which had a higher expression in the bat hand than the foot at both developmental stages (CS17 FC = 1.3; CS16 FC = 1.3). This gene is required for limb induction in the developing mouse, by mediating the reciprocal regulation loop between FGF8 and FGF10 (Xu et al., 1998). In a model proposed by Weatherbee et al. (2006), the relatively lowered expression of this gene in the bat foot plate may play a role in reducing the *Fgf* signalling in the bat foot plate and so increase the incidence of cell death in this tissue.

In contrast to the low level of *Fgf8* expression, *Fgf9* had much higher signal in the CS17 bat hand plate as compared to the mouse (FC = 2.1). In contrast, *Fgf18* had higher signal in the mouse hand plate (FC = 1.4). The signal between the bat hand and foot plates of both stages was not significantly different for either of these genes. These two FGF ligands are known to regulate skeletal development (Hung et al., 2007). *Fgf9* is localised in the mesenchyme surrounding the skeletal condensations at E12.5 while *Fgf18* is expressed in the perichondrium of the developing skeletal elements. Both ligands act to promote chondrocyte proliferation and differentiation. Mutants lacking *Fgf9* expression have shortened stylopods while those lacking this *Fgf18* exhibit delayed ossification (Hung et al., 2007; Liu et al., 2007). The overexpression of the one ligand and the underexpression of the other may provide a novel means of regulating the skeletal elements within the bat. However, an additional mechanism must be responsible for the differences found between the bat hand and foot plates as these genes are similarly expressed in both. This could either be through the restriction of the expression domains to specific regions in each autopod or through the differential regulation of downstream effectors.

Along with these new bat limb candidate genes that were identified in the *Fgf* gene family, two genes in the *Bmp* family were found to have signal that justified further investigation. *Bmp10* showed a much lower expression in the mouse than in the bat hand plate (FC = 3.1), but there were no expression differences between the bat hand and foot plate of either stage. This gene has been shown to be involved in heart development with deletion mutations causing a reduction in cell proliferation (Chen et al., 2004). Therefore, *Bmp10* may be involved in the initiation or maintenance of novel proliferation events in the developing bat hand. It has been suggested that this gene may act through the BMP receptor *Bmpr1a* (Yang et al., 2006). The expression of this receptor did not differ among any of the autopod conditions in this microarray study. However, *Bmpr1b,* had a significantly higher expression in the bat than the mouse (FC = 6.8). This overexpression is also seen in the CS16 HP as compared to the FP (FC = 1.4). *Bmpr1b* is expressed in precartilaginous condensations and serves as the major transducer of BMP signalling in these condensations, promoting chondrogenesis (Yoon and Lyons, 2004; Yoon et al., 2005). It has also been shown to be expressed in the distal region of the chick digit condensations, in a zone known as the phalanx forming region (PFR) (Suzuki et al., 2008). This region commits cells to the chondrogenic condensations and the gene expression in this region has been correlated with digit identity (Suzuki et al., 2008). Mice with mutations in *Bmpr1b* show reduced condensation due to apoptosis and decreased proliferation resulting in shortened phalanges (Baur et al., 2000; Yi et al., 2000). The differential expression found in this microarray study indicates that *Bmpr1b* plays a role in the elongation of these skeletal elements in the bat hand. Different levels of this gene may occur in the PFR of the individual bat digits giving each phalange a unique identity and morphological fate to those of the mouse or the bat hindlimb.

4.4.6 The expression of *Prrx1* supports previous bat limb development studies

Another recent publication in the field of bat limb development examined the *Paired-related homeobox* gene (*Prrx1*, also known as *Prx1*), which is known to promote limb skeletal elongation (Cretekos et al., 2008). This gene was shown to have a different expression pattern between the mouse and the bat hand plate. *Prrx1* was strongly expressed in the CS16 bat hand and had no expression in the mouse E12.5 hand. At CS17 the expression changed and become localised in the perichondrium of the skeletal elements. These differences were found to be driven by changes in the *cis*-regulatory regions that control *Prrx1* expression.

The *Prrx1* gene has been shown to be promote limb skeletal elongation in the bat, with different gene expression patterns occurring between the hand plate of the bat and that of the mouse (Cretekos et al., 2008). These findings were supported by the microarray data on *Prrx1* expression, which was differentially expressed both between the CS17 bat hand and the mouse hand (FC = 1.8) and between the CS17 bat hand and foot (FC = 1.4). Though not significant, *Prrx1* also had a higher expression in the CS16 bat hand as compared to the foot. These expression differences were not related to the expression of its upstream transcription factors as the *Prrx1* transcriptional activator, *Ptk2* did not show any differential expression. *Tenascin-C* (*Tnc)*, a downstream target of PRRX1 (Jones et al., 2001; McKean et al., 2003), showed a very similar pattern of differential expression to that of *Prrx1*. This indicates that the overexpression of *Prrx1* affects the regulation of downstream processes. The expression of *Prrx2*, which has a redundant function to that of *Prrx1* (ten Berge et al., 1998), did not differ between any of the autopod comparisons. The fact that expression is reduced both in the hand and in the foot provides a strong case for the involvement of this gene in specifying the different limb type morphologies. However, this feature also indicates that other factors, in addition to the *cis*-regulatory sequence modification, must be involved to make this up-regulation hand plate specific. Though *Prrx1* remains a good candidate gene, further work needs to be done in explaining how expression differences occur within the bat limb types in addition to those that occur between the mouse and the bat.

4.4.7 Examination of *Meis2* expression in the developing limb

The most robust and intriguing differential gene expression in this study was for a 5′-*Meis2* probe which had significantly higher expression in the bat CS17 HP than in the mouse HP (FC = 13.1) and was also higher in the bat HP than the FP for both stages (CS17 FC = 7.1; CS16 = 2.2). This expression, together with the fact that MEIS2 has been shown to be involved in limb developmental processes (Mercader et al., 2005), make this gene a relevant potential candidate for further investigation in bat hand development.

Meis2 (in conjunction with *Meis1*) codes for a transcription factor that specifies the P-D axis during limb development (Mercader et al., 2000; Mercader et al., 2005). MEIS1, 2 and 3 are members of the TALE family of homeodomain proteins. They all have a three-amino-acid loop extension between homeodomain helices 1 and 2 which is near the C-terminal of the protein (Bürglin, 1997). They also have a conserved domain closer to the N- terminal which allows them to interact with members of another TALE family, *pre-B-cell leukaemia*

transcription factor (*Pbx) 1*, *2*, *3* and *4*. (Knoepfler et al., 1997). The *Meis* genes are important regulators of PBX activity, which is involved in a variety of developmental processes such as cytoskeleton assembly regulation, transcription regulation, signal transduction, cell adhesion and extracellular matrix interaction, (see Laurent et al., (2008) for a review). The MEIS proteins serve to bind to PBX proteins through their conserved N-terminal domains and transport them into the nucleus. This heterodimeric complex then binds to DNA, through the conserved homeodomain region of both molecules, and initiates transcription (Knoepfler et al., 1997). MEIS has also been shown to form trimeric complexes with DNA bound PBX and HOX proteins, stabilising their interaction and promoting transcription (Shanmugam et al., 1999).

During limb initiation and outgrowth the expression of these *Meis* genes is activated in the proximal region of the limb bud by retinoic acid (RA) signalling (Mercader et al., 2000). Retionic acid is synthesised in the lateral plate mesoderm and diffuses in a P-D gradient along the limb bud (Mic et al., 2004). This gradient has been shown to be maintained through the expression of *Cyp26b1* (a gene coding for an enzyme that breaks down RA) in the distal region of the limb bud (Yashiro et al., 2004). However it has been suggested that *Fgf* (*Fgf8*, *Fgf4*, *Fgf9* and *Fgf17*) expression in the AER may also be involved in this distal antagonism (Mariani et al., 2008; Mercader et al., 2000). The interactions of these genes give the cells in the developing limbs their P-D identity (Mercader et al., 2005). Later in development, (E13.5) a second phase of *Meis1*, *2* and *3* expression is initiated by RA signalling in the interdigital regions of the autopods, just prior to apoptosis (Oulad-Abdelghani et al., 1997; Visel et al., 2004). It is unknown what function the expression of this gene may play in this region. However, known interactions of MEIS with several *Hox* gene products suggest that *Meis2* expression may be involved in autopod patterning and differentiation events (Shanmugam et al., 1999). *Meis2* was not differentially expressed between the hand and foot plate of E12.5 mice (Shou et al., 2005), however it was over three fold up regulated in both the fore- and hindlimb of E11.5 mice exposed to RA (Qin et al., 2002). RA signalling and subsequent *Meis* expression in the interdigital regions of the autopods is known to precede apoptotic events, therefore any increase in these levels would promote apoptosis. Interdigital regression due to apoptotic events does not occur in the bat hand plates, however many of the genes known to be involved in this process are still being expressed (Weatherbee et al., 2006).

The conserved domains of the MEIS2 protein play a vital role in its function. However, the modular nature of these domains means that the loss of one would provide the protein with a novel function. This can be seen in the mouse a transcript variant of *Meis2* which has a novel reading frame. This transcript codes for a protein that loses part of its C-terminal homeodomain. This truncated protein can not bind to its specific DNA regulatory sequence and instead acts as a repressor of transcription by other MEIS2 proteins (Yang et al., 2000). These homeodomain-less proteins have been found in the MEIS homologues of different species. The *Drosophila* homologue of *Meis*, *Homothorax* (*Hth*) was found to code for two protein isoforms, one of which lacked a functional homeodomain. These two isoforms had different localisations an functions during development (Noro et al., 2006). The corresponding isoforms were then isolated from the mouse *Meis1* showing that the truncation of the homeodomain was a conserved mechanism for regulating transcription during development (Noro et al., 2006). The examination KNOX transcriptional regulators, genes related to *Meis*, in *Arabidopsis*, revealed a gene that lacked a functional homeodomain (Magnani and Hake, 2008). This gene, *Knatm*, was expressed in the proximal-lateral domains of organ primordia and was identified as a transcriptional regulator involved in P-D patterning (Magnani and Hake, 2008). The type 5′ *Meis2* transcript that was identified as being overexpressed in the bat hand plate is predicted to lack a homeodomain coding region. The resulting protein would therefore not be able to bind to its specific DNA regulatory sequence. However, its ability to bind to its cofactors would be retained and therefore it may act as a competitive inhibitor for transcription, as reported in other studies (Magnani and Hake, 2008; Noro et al., 2006; Yang et al., 2000). The overexpression of *Meis2* has been shown to be related to limb developmental defects associated with P-D reductions (Mercader et al., 2005; Qin et al., 2002). Therefore the overexpression of a negative regulator of MEIS activity may promote elongation events in the developing bat hand plate.

4.4.8 Application of microarray technology for CSH

There has been much discussion about CSH application, experimental procedures and data pre-processing (Bar-Or et al., 2007a). However, relatively little has been said about which normalisation approaches should be used in these experiments. Normalisation transformations are an integral part of the microarray experiment as these steps correct for technical variation. It is a potentially hazardous process if applied indiscriminately as the strategy that is chosen will have a significant effect on the results obtained (Hoffmann et al., 2002). It is important that the assumptions of normalisation methods chosen are met by the

constraints of the experimental design. It is also important to identify the sources of technical variation in a microarray data set and to systematically minimise them.

The reference design used in this experiment (with the test conditions comprising of four different bat autopod samples and the common reference consisting of a pool of mouse hand plates) was similar to that of a CSH study on the development of beak morphology in different species of Darwin's finches (Abzhanov et al., 2006). It was assumed that the majority of the genes would have similar expression in the mouse E13.5 hand plate and the bat hand and foot plates at both CS17 and CS16. This fundamental assumption was informed by previous work on limb development (Margulies et al., 2001; Shou et al., 2005). Robust-spline and R-quantile normalisation procedures based on this assumption succeeded in standardising data, making the overall signal of the probes highly correlated across all conditions. Forcing the data into a distribution whereby the majority of genes are assumed to remain unchanged does pose certain risks, large scale changes between the bat and the mouse may be lost leaving only a subset of differentially expressed genes from which to elucidate different biological processes. However, it is considered a necessary assumption to make to remove technical variation inherent in microarray data sets.

The technical variations found in the dataset of this experiment included within array variation (dye bias, spatial effects) and between array variations (scanning differences, batch effects). Though biases were initially seen in the incorporation rates between the Cy3 (green) and Cy5 (red) fluorescent dyes, this was expected and the final sample concentrations of these two dyes did not differ prior to hybridisation. Biases seen in the fluorescence of the dyes were attributed to photo-bleaching and were compensated for by balancing the channels when scanning. However, high background fluorescence of the Cy3 (green) dye necessitated the use of individual channel correction to reduce the signal of the green channel while maintaining that of the red channel. An additional result of this correction was to slightly reduce batch effects that occurred between arrays due to the high prevalence of low intensity spots which are susceptible to variations in the background found between the channels. Robust-spline within array normalisation worked well in centring the distribution of the log ratios on zero. It dealt well with the high number of low intensity spots found on the arrays, thereby reducing the intensity dependant dye biases found in certain batches. These methods were found to be sufficient in dealing with within array variations, allowing the relative signal

of the different channels and the different probes within one array to be compared accurately. To do the same for the probe set signals across different arrays (both within and between batches) the R-quantile method of between array normalisation was performed. The assumptions of this method, that all the red channel values are the same, was met by using the same mouse hand plate pool as the reference in each experiment. The red channels were considered to be technical replicates and were forced into the same distribution. The green channels were simultaneously transformed by anchoring the log-ratios, thereby maintaining their inherent biological variability. This method allowed the single channel data to be analysed but had no impact on the log-ratio values. Due to this, batch effects found between the log-ratios of different arrays had to be removed using a specific batch correction function, this allowed the low level differential expression that occurred between the intraspecies comparisons to be identified.

There is no robust filtering function to remove poorly hybridised probes in CSH microarrays (Bar-Or et al., 2007a). In this experiment a comprehensive filtering process was implemented where the genes with low, poor quality or consistent spot signal across all experiments, were eliminated. Roughly 35% of the array was retained as having signal suitable for analysis. This was comparable to previous studies using commercial arrays for CSH (Abzhanov et al., 2006; Bar-Or et al., 2006). It was also much higher than the signal (15%) obtained when hybridising bat skeletal muscle transcripts to a human cDNA platform (Eddy and Storey, 2001). Overall, the cross-species hybridisation of bat transcripts to mouse specific probes was considered successful, resulting in strong signal from a subset of genes in the bat limb conditions which were associated to gene annotations that were not significantly different from that of the mouse. However, three genes known to be differentially expressed between the fore- and hindlimb of the mouse (*Tbx4*, *5* and *Pitx1*) did not follow the same trends in the bat. It would be informative to know the sequences and validate the expression levels of these genes in *M. natalensis*. The mRNA of these genes could be amplified using degenerate primers, cloned and then sequenced. Once the sequences are known, qRT-PCR could be performed using specifically designed primer and accurate measures of expression for each gene obtained. This would indicate whether the ambiguous results from the microarray data were due to sequence dissimilarity, biological differences or poor resolution of the dataset.

The interspecies comparison between the bat hand plate and the mouse hand plate exhibited high fold changes and had signal correlations that were lower than that exhibited by the intraspecies comparisons. These comparisons were highly correlated and reflected the more typical relationship between developing autopod transcriptome experiments (Shou et al., 2005). The low prevalence of high fold change differences in the intraspecies comparisons indicated that these high fold changes between the interspecies comparisons should be treated with caution. Over 80% of these high fold-change genes were similarly up or down regulated among all the bat limb conditions, indicating that differences between the bat autopods and the mouse hand plate were not reflected in the bat autopod comparisons. This suggests that this effect is not due to the slight differences in stages between the bat and the mouse. Nor is it due to sequence dissimilarity, as this would have resulted in a loss of signal from the bat autopod samples, instead a distinct subset of genes shown to have higher signal in the bat autopod samples. It was therefore attributed to either, dye biases that remain uncorrected by normalisation or biological species-specific differences in genes that may not be directly involved in specifying the distinct limb morphologies. These factors can potentially confound attempts to isolate candidate genes. This was avoided by using the multiple comparisons available in this experiment as biological filters. These allow significantly differentially expressed genes that are common in two or more comparisons to be isolated and, either excluded from, or included in a more stringent gene list. This method has been used to isolate genes that were commonly regulated in both the autopod and the genitals of the mouse (Cobb and Duboule, 2005). A novel application of this method was to use the gene list redundancy between two different interspecies analyses to isolate and exclude genes that were commonly differentially expressed among the different tissues in the one species. This method was then used to isolate genes that were commonly regulated between two stages of bat development. Gene lists produced using biological filtering were more robust due to the exclusion of both false positives and false negatives. They were also more informative, as candidate genes were picked based on their prolonged over expression across two stages of development. In this experiment potential pitfalls in the CSH hybridisation bias were removed through the careful use of an appropriate normalisation and pre-processing pipeline and the accumulation of false positive and negative signals was avoided by the stringent cross validation of differential expression among lists of differentially expressed genes. This resulted in a high quality data set that could be used to analyse the differences in gene expression between the morphologically different autopods of the bat and the mouse.

Chapter 5

Conclusions and Recommendations

5.1 THE FUTURE OF CROSS SPECIES MICROARRAYS

This study validates the use of microarray analysis in evolutionary studies. The cross-species hybridisation of bat transcripts to mouse specific probes was considered successful. Though there were issues with low and/or unbalanced signal, conventional normalisation procedures were able to transform the dataset into one that was more amenable to further analysis. The refinement of this dataset into a smaller subset of probes through the use of a novel series of post normalisation filtering steps increased the power of the differential analysis and allowed genes to be identified under stringent statistical conditions. The use of several analysis techniques in addition to a novel biological filtering method ensured that the genes isolated were robust and suitable for further analysis. This experiment provides an additional case study in the growing field of cross species microarrays and should help in motivating the use of this technique in other research enterprises. This cross species microarray illustrates how non-model organisms can be used as genetic resources and may promote interest in designing arrays that are specific to these organisms. These would provide evolutionary and developmental biologists with even more powerful and accurate tools than are currently available. This cross species microarray analyses provided an efficient method of acquiring a large amount of data in a short space of time, and should populate the field of bat limb development with additional candidate genes to explore and hopefully encourage new avenues of research. Based on the results of this study, cross species microarray analyses are recommended as a powerful technique for identifying candidate genes in an unbiased manner. Some of the genes isolated in this study were well characterised genes that were known to have important roles in limb development (i.e. the 5′ *Hoxd* genes). Many were genes which, though known to play a role in limb development, may never have been chosen as potential candidates for research in bat limb development due to their relatively low profile in the literature. Others were, and still are, unknowns, and may point towards new and exciting directions to be taken in the field of limb development research.

5.2 VERIFICATION OF MICROARRAY DATA

In this study 50 new candidate genes which may play a role in shaping the unique bat wing were identified. Many of these genes are known from previous to work to play a role in developmental patterning, cell proliferation and differentiation or apoptosis indicating that they may be of interest in bat limb development. However microarray analyses are not definitive. Due to the large numbers of genes being tested, the low level of replication and the inherent variability of the dataset, these results must always be validated (Chuaqui et al., 2002). This is conventionally done using two different methods, Quantitative Reverse Transcriptase Polymerase Chain Reaction (*q*RT-PCR) and *in situ* hybridisation. The *q*RT-PCR method provides a quantitative measure of the abundance of a particular transcript in the samples being tested. Data such as fold change information or relative abundance of transcripts would be directly comparable between these experiments and that of the microarray analysis. In *in situ* hybridisation, the tissue being examined is probed directly to measure the spatial and temporal patterns of expression of the transcript of interest.

5.3 POSSIBLE ROLE OF CANDIDATE GENES IN LIMB DEVELOPMENT AND FUTURE STUDIES

Examination of groups of genes belonging to families or pathways of interest in limb development confirmed the results of many of the studies looking at these genes in bat limb development, including the well characterised *Hoxd13*. It is suggested that the expression of all of these genes along with that of *Hoxa11* should be examined in the bat autopods. Further work should also be done on the genes downstream of *Hoxd* gene expression in bat autopods, some of these genes have been identified in a gene expression study on $Hoxd^{Del1\text{-}13}$ mutant mouse limbs (Cobb and Duboule, 2005). On the basis of the microarray data presented in this study, they present them selves as promising candidate genes to be examined in the context of bat limb development,

An examination of the bat GCR illustrated that sequence changes in this enhancer regions lead to novel expression domains of the *Hoxd* genes in transgenic mice (Ray and Capecchi, 2008). This work should be complemented by the examination of the bat *Prox* sequence, including its conserved sequence CsC. This sequence should be isolated in several bat species and compared to that of the mouse and other tetrapods. Any changes in conserved regions of these regulatory elements may play a role in changing the transcriptional activation

of the 5′ *Hoxd* genes in the bat. These changes may lead to different temporal and spatial patterns of expression of the genes of the *Hoxd* cluster (Gonzalez et al., 2007). In addition to this, the gene expression pattern of all of the 5′ *Hoxd* genes should be examined, both in the bat (Ray and Capecchi, 2008) and in transgenic mice that carry either the bat CGR or the bat *Prox* enhancer regions as well as those that carry both. In this way the enhancer activity domains of these collaborative enhancers can be distinguished. Retinoic acid signalling has also been identified in this study as a possible line of research in bat limb development studies. The expression patterns of the major genes involved in RA synthesis and degradation (particularly *Rdh9,-10, -11 -14* and *Cyp26a1*) should be examined in the autopods to determine whether the regional control of this morphogen has been changed in the bat. Further characterisation of the specific transcript variant that is responsible for the signal from the 3′-*Meis2* probe is needed. It may be that the protein product of this novel splice variant would perform a regulatory role in the interactions between MEIS, PBX and HOX proteins, as has been shown in the *Meis2* homologues of other species (Human *Meis2*: (Yang et al., 2000) ; Common Fruit fly - *Hth*: (Noro et al., 2006) ; *Arabidopsis Knox*: (Magnani and Hake, 2008)). The presence of this alternatively spliced *Meis2* transcript remains to be identified in the bat, and the possible role that a truncated protein product may play in limb development should be explored. This would allow regulatory functions of different isoforms of MEIS2 to be elucidated. It would be interesting to explore whether altered RA levels play a role in the alternative splicing of *Meis2*.

The above three lines of research are not mutually exclusive. Examination of the metabolism and signalling of RA in bat limbs should be done in conjunction with studies on expression of the 5′-*Hoxd* and *Meis*2 gene expressions, and possible interaction between these genes and pathways should be explored. This would enable a model, integrating RA signalling, *Meis2* expression and possible MEIS2 and HOX interactions in the context of bat limb development, to be developed. This model could be tested using ectopic expression studies in the chick. Further information on possible novel bat gene sequences or regulatory regions could also be gained through transgenic experiments in the developing mouse. Additional work should be done at the protein level, to gain insight about the actual abundance and possible actions and interactions of these proteins. This would provide a holistic view of the expression and regulation of primary limb developmental genes in the context of autopod development in the bat as opposed to the mouse.

5.4 FURTHER ANALYSIS OF DATASET

The analyses performed on the microarray data in this study are not exhaustive. A conservative approach was taken which allowed only robust genes that passed stringent statistical criteria (multiple testing adjustments) to be classified as significantly differentially expressed. This served to create several short lists of robust candidate genes for further analysis. With additional data mining techniques more information could be extracted from this dataset to create more populated gene lists for further functional analysis. This would aid in creating a broader view of what is happening in the developing limb instead of focusing on the functions of a handful of genes. The real power of this microarray dataset lies in being able to obtain a 'big picture' of what is happening in the limb at any one time. The challenge lies in being able to interpret the data from this perspective as there is a confounding amount of information available. This process is aided by powerful software that allows for the integration of the microarray data into extensive gene pathway databases. This data would benefit from further exploration using such technologies. It is hoped that this dataset will continue to provide valid and relevant information on the molecular processes that are occurring to specify the bat wing morphology during development.

Chapter 6

List of References

Abzhanov, A., Kuo, W.P., Hartmann, C., Grant, B.R., Grant, P.R. and Tabin, C.J. (2006). The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature* **442**, 563-567.

Acharya, C.R., Hsu, D.S., Anders, C.K., Anguiano, A., Salter, K.H., Walters, K.S., Redman, R.C., Tuchman, S.A., Moylan, C.A., Mukherjee, S. et al. (2008). Gene expression signatures, clinicopathological features, and individualized therapy in breast cancer. *JAMA* **299**, 1574-1587.

Adams, R.A. (1992a). Comparative skeletogenesis of the forearm of the little brown bat (*Myotis lucifungus*) and the Norway rat (*Rattus norvegicus*). *Journal of Morphology* **214**, 251-260.

Adams, R.A. (1992b). Stages of development and sequence of bone formation in the little brown bat, *Myotis lucifungus*. *Journal of Mammalogy* **73**, 160-167.

Adjaye, J., Herwig, R., Herrmann, D., Wruck, W., BenKahla, A., Brink, T.C., Nowak, M., Carnwath, J.W., Hultschig, C., Niemann, H. et al. (2004). Cross-species hybridisation of human and bovine orthologous genes on high density cDNA microarrays. *BMC Genomics* **5**.

Al-Shahrour, F., Minguez, P., Tárraga, J., Montaner, D., Alloza, E., Vaquerizas, J.M., Conde, L., Blaschke, C., Vera, J. and Dopazo, J. (2006). BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. *Nucleic Acids Research* **34**, W472-W476.

Altringham, J.D. (1996). Bats: Biology and behaviour. Oxford: Oxford University Press.

American Veterinary Medical Association. (2001). 2000 Report of the AVMA panel on euthanasia. *Journal of American Veterinary Medical Association* **218**, 669-696.

Anders, C.K., Hsu, D.S., Broadwater, G., Acharya, C.R., Foekens, J.A., Zhang, Y., Wang, Y., Marcom, P.K., Marks, J.R., Febbo, P.G. et al. (2008). Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. *Journal of Clinical Oncology* **26**, 3324-3330.

Axon Instruments Inc. (2004). GenePix® Pro 6.0 Microarray aquisition and analysis software for Genepix microarray scanners user's guide & tutorial. *Axon Instruments Inc. User Guide* **Part Number 2500-0137 Rev K**.

Bar-Or, C., Bar-Eyal, M., Gal, T.Z., Kapulnik, Y., Czosnek, H. and Koltai, H. (2006). Derivation of species-specific hybridization-like knowledge out of cross-species hybridization results. *BMC Genomics* **7**.

Bar-Or, C., Czosnek, H. and Koltai, H. (2007a). Cross-species microarray hybridizations: a developing tool for studying species diversity. *TRENDS in Genetics* **23**, 200-207.

Bar-Or, C., Novikov, E., Reiner, A., Czosnek, H. and Koltai, H. (2007b). Utilizing microarray spot characteristics to improve cross-species hybridization results. *Genomics* **90**, 636-645.

Baur, S.T., Mai, J.J. and Dymecki, S.M. (2000). Combinatorial signaling through BMP receptor IB and GDF5: shaping of the distal mouse limb and the genetics of distal limb diversity. *Development* **127**, 605-619.

Bengtsson, A. (2003). Microarray image analysis: Background estimation using region and filtering techniques. In *Centre for Mathematical Sciences*, vol. MSc. Sweden: Lund Institute of Technology.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B, Statistical methodology* **57**, 289-300.

Bernard, R.T.F. (1980). Reproductive cycles of *Miniopterus schreibersi natalensis* (Kuhl, 1819) and *Miniopterus fraterculus* (Thomas and Schwann, 1906). *Annals of the Transvaal Museum* **32**, 55-64.

Bernard, R.T.F. (1994). Reproductive synchrony and annual variation in foetal growth rate in the long-fingered bat (*Miniopterus schreibersii*). *Journal of Zoology: Proceedings of the Zoological Society of London* **232**, 485-490.

Bernard, R.T.F., Bojarski, C. and Millar, R.P. (1991). Plasma progesterone and luteinizing hormone concentrations and the role of the corpus luteum and LH gonadotrophs in the control of delayed implantation in Schreibers' long-fingered bat (*Miniopterus schreibersii*). *Journal of Reproduction and Fertility* **93**, 31-42.

Bernard, R.T.F., Cotterill, F.P.D. and Fergusson, R.A. (1996). On the occurrence of a short period of delayed implantation in Schreibers' long-fingered bat (*Miniopterus*

schreibersii) from a tropical latitude in Zimbabwe. *Journal of Zoology: Proceedings of the Zoological Society of London* **238**, 13-22.

Bininda-Emonds, O.R.P., Jeffery, J.E., Sánchez-Villagra, M.R., Hanken, J., Colbert, M., Pieau, C., Selwood, L., ten Cate, C., Raynaud, A., Osabutey, C.K. et al. (2007). Forelimb-hindlimb developmental timing changes across tetrapod phylogeny. *BMC Evolutionary Biology* **7**.

Birney, E., Andrews, T.D., Bevan, P., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cuff, J., Curwen, V., Cutts, T. et al. (2004). An overview of Ensembl. *Genome Research* **14**, 925-928.

Boulet, A.M. and Capecchi, M.R. (2002). Duplication of the *Hoxd11* gene causes alterations in the axial and appendicular skeleton of the mouse. *Developmental Biology* **249**, 96-107.

Boulet, A.M. and Capecchi, M.R. (2004). Multiple roles of *Hoxa11* and *Hoxd11* in the formation of the mammalian forelimb zeugopod. *Development* **131**, 299-309.

Boutros, P. *Bioconductor Forum - Increasing rlm iterations for convergence in limma's normalizeRobustSpline* [Online]. Available: https://stat.ethz.ch/pipermail/bioconductor/2004-August/005641.html. June 2008

Boylan, J.F. and Gudas, L.J. (1992). The level of CRABP-I expression influences the amounts and types of all- trans-retinoic acid metabolites in F9 teratocarcinoma stem cells. *Journal of Biological Chemistry* **267**, 21486-21491.

Brodsky, L.I., Jacob-Hirsch, J., Avivi, A., Trakhtenbrot, L., Zeligson, S., Amariglio, N., Paz, A., Korol, A.B., Band, M., Rechavi, G. et al. (2005). Evolutionary regulation of the blind subterranean mole rat, *Spalax*, revealed by genome-wide gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 17047-17052.

Brown, C.R. and Bernard, R.T.F. (1994). Thermal preference of Schreiber's long-fingered (*Miniopterus schreiberisii*) and Cape horseshoe (*Rhinolophus capensis*) bats. *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology* **107**, 439-449.

Brown, P.O. and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nature Genetics Supplement* **21**, 33-37.

Buchanan, G.D. and Younglai, E.V. (1986). Plasma progesterone levels during pregnancy in the little brown bat *Myotis lucifugus* (Vespertilionidae). *Biology of Reproduction* **34**, 878-884.

Bürglin, T.R. (1997). Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic Acids Research* **25**, 4173-4180.

Burns, J.M. and Easley, R.G. (1977). Hormonal control of delayed development in the California leaf-nosed bat, *Macrotus californicus*: III. Changes in plasma progesterone during pregnancy. *General Comparative Endocrinology* **32**, 163-166.

Capellini, T.D., Di Giacomo, G.S., Valentina, Brendolan, A., Ferretti, E., Srivastava, D., Zappavigna, V. and Selleri, L. (2006). *Pbx1/Pbx2* requirement for distal limb patterning is mediated by the hierarchical control of Hox gene spatial distribution and *Shh* expression. *Development* **133**, 2263-2273.

Chaudhry, A.Z., Lyons, G.E. and Gronostajski, R.M. (1997). Expression patterns of the four nuclear factor I genes during mouse embryogenesis indicate a potential role in development. *developmental Dynamics* **208**, 313-325.

Chen, C.-H., Cretekos, C.J., Rasweiler, J.J., IV and Behringer, R.R. (2005). *Hoxd13* expression in the developing limbs of the short-tailed fruit bat, *Carollia perspicillata*. *Evolution & Development* **7**, 130-141.

Chen, H., Shi, S., Acosta, L., Li, W., Lu, J., Bao, S., Chen, Z., Yang, Z., Schneider, M.D., Chien, K.R. et al. (2004). BMP10 is essential for maintaining cardiac growth during murine cardiogenesis. *Development* **131**, 2219-2231.

Chen, Y.P. and Zhao, X. (1998). Shaping limbs by apoptosis. *Journal of Experimental Zoology Part A: Comparative Experimental Biology* **282**, 691-702.

Chiu, C.-H. and Hamrick, M.W. (2002). Evolution and development of the primate limb skeleton. *Evolutionary Anthropology* **11**, 94-107.

Chuaqui, R.F., Bonner, R.F., Best, C.J.M., Gillespie, J.W., Flaig, M.J., Hewitt, S.M., Phillips, J.L., Krizman, D.B., Tangrea, M.A., Ahram, M. et al. (2002). Post-analysis follow-up and validation of microarray experiments. *Nature Genetics* **32**, 509-514.

Churchill, G.A. (2002). Fundamentals of experimental design for cDNA microarrays. *Nature Genetics* **32**, 490-495.

Cobb, J. and Duboule, D. (2005). Comparative analysis of genes downstream of the Hoxd cluster in developing digits and external genitalia. *Development* **132**, 3055-3067.

Cohen, R., Chalifa-Caspi, V., Williams, T.D., Auslander, M., George, S.G., Chipman, J.K. and Tom, M. (2007). Estimating the efficiency of fish cross-species cDNA microarray hybridization. *Marine Biotechnology* **9**, 491-499.

Cox, W.G., Beaudet, M.P., Agnew, J.Y. and Ruth, J.L. (2004). Possible sources of dye-related signal correlation bias in two-color DNA microarray assays. *Analytical Biochemistry* **331**, 243-254.

Cretekos, C.J., Deng, J.-M., Green, E.D., NISC Comparative Sequencing Program, Rasweiler, J.J., IV and Behringer, R.R. (2007). Isolation, genomic structure and developmental expression of *Fgf8* in the short-tailed fruit bat, *Carollia perspicillata*. *International Journal of Developmental Biology* **51**, 333-338.

Cretekos, C.J., Rasweiler, J.J., IV and Behringer, R.R. (2001). Comparative studies on limb morphogenesis in mice and bats: A functional genetic approach towards a molecular understanding of diversity in organ formation. *Reproduction, Fertility and Development* **13**, 691-695.

Cretekos, C.J., Wang, Y., Green, E.D., NISC Comparative Sequencing Program, Martin, J.F., Rasweiler IV, J.J. and Behringer, R.R. (2008). Regulatory divergence modifies limb length between mammals. *Genes & Development* **22**, 141-151.

Cretekos, C.J., Weatherbee, S.D., Chen, C.-H., Badwaik, N.K., Niswander, L., Behringer, R.R. and Rasweiler, J.J.I. (2005). Embryonic staging system for the Short-Tailed Fruit Bat, *Carollia perspicillata*, a model organism for the mammalian order *Chiroptera*, based upon timed pregnancies in captive-bred animals. *developmental Dynamics* **233**, 721-738.

Crichton, E.G., Seamark, R.F. and Krutzsch, P.H. (1989). The status of the corpus luteum during pregnancy in *Miniopterus schreibersii* (Chiroptera: Vespertilionidae) with emphasis on its role in developmental delay. *Cell and Tissue Research* **258**, 183-201.

Cui, X. and Churchill, G.A. (2003). Statistical tests for differential expression in cDNA microarray experiments. *Genome Biology* **4**.

Cumming, G.S. and Bernard, R.T.F. (1997). Rainfall, food abundance and timing of parturition in African bats. *Oecologia* **111**, 309-317.

Currie, W.B., Blake, M. and Wimsatt, W.A. (1988). Fetal development, and placental and maternal plasma concentrations of progesterone in the little brown bat (*Myotis lucifugus*). *Journal of Reproduction and Fertility* **82**, 401-407.

Darwin, C. (1890). The origin of the species by means of natural selection. London: John Murray.

Davis, A.P., Witte, D.P., Hsieh-Li, H.M., Potter, S.S. and Capecchi, M.R. (1995). Absence of radius and ulna in mice lacking *hoxa-11* and *hoxd-11*. *Nature* **375**, 791-795.

Diehl, K.-H., Hull, R., Morton, D.B., Pfister, R., Rabemampianina, Y., Smith, D., Vidal, J.-M. and van de Vorstenbosch, C. (2001). A good practice guide to the administration of substances and removal of blood, including routes and volumes. *Journal of Applied Toxicology* **21**, 15-23.

Ding, Y., Xu, L., Chen, S., Jovanovic, B.D., Helenowski, I.B., Kelly, D.L., Catalona, W.J., Yang, X.J., Pins, M., Ananthanarayanan, V. et al. (2006). Characterization of a method for profiling gene expression in cells recovered from intact human prostate tissue using RNA linear amplification. *Prostate Cancer and Prostatic Diseases* **9**, 379-391.

Dünker, N., Schmitt, K. and Krieglstein, K. (2002). TGF-β is required for programmed cell death in interdigital webs of the developing mouse limb. *Mechanisms of Development* **113**, 111-120.

Eddy, S.F. and Storey, K.B. (2001). Gene expression in hibernation: Testing skeletal muscle of little brown bats, *Myotis Lucifugus*, using commercially available cDNA microarrays. In *Proceedings of the Virtual Conference in Genomics and Bioinformatics*.

Eddy, S.F. and Storey, K.B. (2002). Dynamic use of cDNA arrays: Heterologous probing for the gene discovery and exploration of organismal adaptation to environmental stress. Amsterdam: Elsevier Press.

Eick, G.N., Jacobs, D.S. and Matthee, C.A. (2005). A nuclear DNA phylogenetic perspective on the evolution of echolocation and historical biogeography of extant bats (Chiroptera). *Molecular Biology and Evolution.* **22**, 1869-1886.

Emmanouil-Nikoloussi, E.-N., Goret-Nicaise, M., Foroglou, P., Thliveris, J.A. and Kerameos-Foroglou, C. (2008). All-trans-retinoic acid-induced disturbance of forelimb digital apoptosis in mouse embryos: a preliminary scanning electron microscope (SEM) tudy. *European Journal of Anatomy* **12**, 25-32.

Farnum, C.E., Tinsley, M. and Hermanson, J.W. (2008a). Forelimb versus hindlimb skeletal development in the big brown bat, *Eptesicus fuscus*: Functional divergence is reflected in chondrocytic performance in autopodial growth plates. *Cells Tissues Organs* **187**, 35-47.

Farnum, C.E., Tinsley, M. and Hermanson, J.W. (2008b). Postnatal bone elongation of the Manus versus Pes: Analysis of the chondrocytic differentiation cascade in *Mus musculus* and *Eptesicus fuscus*. *Cells Tissues Organs* **187**, 48-58.

Findley, J.S., Studier, E.H. and Wilson, D.E. (1972). Morphologic properties of bat wings. *Journal of Mammalogy* **53**, 429-444.

Gentleman, R., Carey, V., Morgan, M. and Falcon, S. (2006). *Biobase*: Base functions for Bioconductor. R package. **Ver. 1.16.1.**

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B.M., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J. et al. (2004). Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology* **5**, R80.

Giannini, N., Goswami, A. and Sánchez-Villagra, M.R. (2006). Development of integumentary structures in Rousettus amplexicaudatus (Mammalia: Chiroptera: Pteropodidae) during late-embryonic and fetal stages. *Journal of Mammalogy* **87**, 993-1001.

Gonzalez, F., Duboule, D. and Spitz, F. (2007). Transgenic analysis of Hoxd gene regulation during digit development. *Developmental Biology* **306**, 847-859.

Gopalakrishna, A. and Chari, G.C. (1983). A review of the taxonomic position of *Miniopterus* based on embryological characters. *Current Science* **52**, 1176-1180.

Gopalakrishna, A. and Karim, K.B. (1979). Fetal membranes and placentation in Chiroptera. *Journal of Reproduction and Fertility* **56**, 417-429.

Gorokhova, E. (2005). Effects of preservation and storage of microcrustaceans in RNA*later* on RNA and DNA degradation. *Limnology and Oceanography: Methods* **3**, 143-148.

Gottardo, R. (2008). *agce*: Analysis of growth curve experiments. R package. **Ver. 1.2.**

Gray, P.A., Fu, H., Luo, P., Zhao, Q., Yu, J., Ferrari, A., Tenzen, T., Yuk, D.I., Tsung, E.F., Cai, Z. et al. (2004). Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science* **306**, 2255-2257.

Grigoryev, D.N., Ma, S.-F., Simon, B.A., Irizarry, R.A., Ye, S.Q. and Garcia, J.G.N. (2005). *In vitro* identification and *in silico* utilization of interspecies sequence similarities using GeneChip® technology. *BMC Genomics* **6**, 62.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignement editor and analysis progrma for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98.

Hastie, T., Tibshirani, R., Narasimhan, B. and Chu, G. (2007). *impute*: Imputation for microarray data. R packages. **Ver. 1.10.0.**

Hayssen, V. and Kunz, T.H. (1996). Allometry of litter mass in bats: Maternal size wing morphology, and phylogeny. *Journal of Mammalogy* **77**, 476-490.

Hill, J.E. and Smith, J.D. (1984). Bats: A natural History. Austin: University of Texas Press.

Hockman, D., Cretekos, C.J., Mason, M.K., Behringer, R.R., Jacobs, D.S. and Illing, N. (2008). A second wave of *Sonic hedgehog* expression during the development of the bat limb. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16982-16987.

Hoffmann, R., Seidl, T. and Dugas, M. (2002). Profound effect of normalization on detection of differentially expressed genes in oligonucleotide microarray data analysis. *Genome Biology* **3**, research0033.1-0033.11.

Hughes, P.M. and Rayner, J.M.V. (1991). Addition of artificial loads to long-eared bats *Plecotus auritus*: Handicapping flight performance. *Journal of Experimental Biology* **161**, 285-298.

Hung, I.H., Yu, K., Lavine, K.J. and Ornitz, D.M. (2007). FGF9 regulates early hypertrophic chondrocyte differentiation and skeletal vascularization in the developing stylopod. *Developmental Biology* **307**, 300-313.

Isaac, S.S. and Marimuthu, G. (1996). Postnatal growth and age estimation in the Indian pygmy bat *Pipistrellus mimus*. *Journal of Mammalogy* **77**, 199-204.

Isogai, Z., Ono, R.N., Ushiro, S., Keene, D.R., Chen, Y., Mazzieri, R., Charbonneau, N.L., Reinhardt, D.P., Rifkin, D.B. and Sakai, L.Y. (2003). Latent Transforming Growth Factor β-binding Protein 1 interacts with Fibrillin and is a microfibril-associated protein. *Journal of Biological Chemistry* **278**, 2750-2757.

Jacobs, D.S. (1999). Intraspecific variation in wingspan and echolocation call flexibility might explain the use of different habitats by the insectivorous bat, *Miniopterus schreibersii* (Vespertilionidae: Miniopteridae). *Acta Chiropterologica* **1**, 93-103.

Jacobs, D.S., Cotterill, F.W., Taylor, P. and Griffin, M. (2004). *Miniopterus natalensis*. In *IUCN 2007. 2007 IUCN red list of threatened species*.

Ji, W., Zhou, W., Gregg, K., Yu, N., Davis, S. and Davis, S. (2004). A method for cross-species gene expression analysis with high-density oligonucleotide arrays. *Nucleic Acids Research* **32**, e93.

Johnson, W.E., Rabinovic, A. and Li, C. (2007). Adjusting batch effects in microarray expression data using Empirical Bayes methods. *Biostatistics* **8**, 118-127.

Jones, C. (1967). Growth, development and wing loading in the evening bat, *Nycticeius humeralis* (Rafinesque). *Journal of Mammalogy* **48**, 1-19.

Jones, F.S., Meech, R., Edelman, D.B., Oakey, R.J. and Jones, P.L. (2001). Prx1 Controls Vascular Smooth Muscle Cell Proliferation and Tenascin-C Expression and Is Upregulated With Prx2 in Pulmonary Vascular Disease. *Circulation Research* **89**, 131-138.

Junker, K., Bain, O. and Boomker, J. (2008). Helminth parasites of Natal long-fingered bats, *Miniopterus natalensis* (Chiroptera: Miniopteridae), in South Africa. *Onderstepoort Journal of Veterinary Research* **75**, 261-265.

Kaufman, M.H. (1992). The atlas of mouse development. London: Academic Press Ltd.

Kerr, M.K. (2003). Design considerations for efficient and effective microarray studies. *Biometrics* **59**, 822-828.

Kimura, K. and Uchida, T.A. (1983). Ultrastructural observations of delayed implantation in the Japanese long-fingered bat, *Miniopterus schreibersii fuliginosus*. *Journal of Reproduction and Fertility* **69**, 187-193.

Knoepfler, P.S., Calvo, K.R., Chen, H., Antonarakis, S.E. and Kamps, M.P. (1997). Meis1 and pKnox1 bind DNA cooperatively with Pbx1 utilizing an interaction surface disrupted in oncoprotein E2a-Pbx1. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 14553-14558.

Kochhar, D.M. (1973). Limb development in mouse embryos. I. Analysis of teratogenic effects of retinoic acid. *Teratology* **7**, 289-298.

Kriegs, J.O., Churakov, G., Kiefmann, M., Jordan, U., Brosius, J. and Schmitz, J. (2006). Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biology* **4**, e91.

Kuhn, A. (2007). *annotationTools*: Annotate microarrays and perform cross-species gene expression analyses using flat file databases. R packages. **Ver. 1.8.0.**

Laurent, A., Bihan, R., Omilli, F., Deschamps, J. and Pellerin, I. (2008). PBX proteins: much more than Hox cofactors. *International Journal of Developmental Biology* **52**, 9-20.

Lawrence, M.A. (1991). Biological Observations on a Collection of New Guinea *Syconycteris australis* (Chiroptera, Pteropodidae) in the American Museum of Natural History. *American Museum Novitates* **3024**, 1-27.

Liu, Z., Lavine, K.J., Hung, I.H. and Ornitz, D.M. (2007). FGF18 is required for early chondrocyte proliferation, hypertrophy andvascular invasion of the growth plate. *Developmental Biology* **302**, 80-91.

Magnani, E. and Hake, S. (2008). *KNOX* Lost the *OX*: The *Arabidopsis KNATM* Gene Defines a Novel Class of KNOX Transcriptional Regulators Missing the Homeodomain. *Plant Cell* **20**, 875-887.

Margulies, E.H., Kardia, S.L.R. and Innis, J.W. (2001). A comparative molecular analysis of developing mouse forelimbs and hindlimbs using serial analysis of gene expression (SAGE). *Genome Research* **11**, 1686-1698.

Mariani, F.V., Ahn, C.P. and Martin, G.R. (2008). Genetic evidence that FGFs have an instructive role in limb proximal–distal patterning. *Nature* **453**, 401-406.

Martínez-Llordella, M., Lozano, J.J., Puig-Pey, I., Orlando, G., Tisone, G., Lerut, J., Carlos Benítez, Jose Antonio Pons, Pascual Parrilla, Pablo Ramírez et al. (2008). Using transcriptional profiling to develop a diagnostic test of operational tolerance in liver transplant recipients. *Journal of Clinical Investigation* **118**.

McCaffery, P. and Simons, C. (2007). Prospective teratology of retinoic acid metabolic blocking agents (RAMBAs) and loss of CYP26 activity. *Current Pharmaceutical Design* **13**, 3020-3037.

McDonald, J.T., Rautenbach, I.L. and Nel, J.A.J. (1990a). Foraging ecology of bats observed at De Hoop Provincial Nature Reserve, southern Cape Province. *South African Journal of Wildlife Research* **20**, 133-145.

McDonald, J.T., Rautenbach, I.L. and Nel, J.A.J. (1990b). Roosting requirements and behaviour of five bat species at De Hoop Guano Cave, southern Cape Province of South Africa. *South African Journal of Wildlife Research* **20**, 157-161.

McKean, D.M., Sisbarro, L., Ilic, D., Kaplan-Alburquerque, N., Nemenoff, R., Weiser-Evans, M., Kern, M.J. and Jones, P.L. (2003). FAK induces expression of Prx1 to

promote enascin-C-dependent fibroblast migration. *Journal of Cell Biology* **161**, 393-402.

McLeod, M.J. (1980). Differential staining of cartilage and bone in whole mouse fetuses by Alcian blue and Alizarin red S. *Teratology* **22**, 299-301.

Mercader, N., Leonardo, E., Piedra, M.E., Martínez-A, C., Ros, M.Á. and Torres, M. (2000). Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. *Development* **127**, 3961-3970.

Mercader, N., Tanaka, E.M. and Torres, M. (2005). Proximodistal identity during vertebrate limb regeneration is regulated by Meis homeodomain proteins. *Development* **132**, 4131-4142.

Mic, F.A., Sirbu, I.O. and Duester, G. (2004). Retinoic acid synthesis controlled by *Raldh2* is required early for limb bud initiation and then later as a proximodistal signal during apical ectodermal ridge formation. *Journal of Biological Chemistry* **279**, 26698-26706.

Miller-Butterworth, C.M., Eick, G., Jacobs, D.S., Schoeman, M.C. and Harley, E.H. (2005). Genetic and phenotypic differences between South African long-fingered bats, with a global miniopterine phylogeny. *Journal of Mammalogy* **86**, 1121-1135.

Miller-Butterworth, C.M., Murphy, W.J., O'Brien, S.J., Jacobs, D.S., Springer, M.S. and Teeling, E.C. (2007). A family matter: Conclusive resolution of the taxonomic position of the long-fingered bats, *Miniopterus*. *Molecular Biology and Evolution.* **24**, 1553-1561.

Mills, M.G.L. and Hes, L. (1997). The complete book of Southern African mammals, pp. 356. Cape Town: Struik.

Montavon, T., Le Garrec, J.-F., Kerszberg, M. and Duboule, D. (2008). Modeling Hox gene regulation in digits: reverse collinearity and the molecular origin of thumbness. *Genes & Development* **22**, 346-359.

Morriss-Kay, G.M. and Sokolova, N. (1996). Embryonic development and pattern formation. *Federation of American Societies for Experimental Biology Journal* **10**, 961-968.

Morton, D.B., Abbot, D., Barclay, R., Close, B.S., Ewbank, R., Gask, D., Heath, M., Mattic, S., Poole, T., Seamer, J. et al. (1993). Removal of blood from laboratory mammals and birds: First Report of the BVA/FRAME/RSPCA/UFAW joint working group on refinement. *Laboratory Animals* **27**, 1-22.

Mouse Genome Informatics Web. *Gene Expression Database (GXD)* [Online]. Available: http://www.informatics.jax.org. January 2009

Murphy, W.J., Eizirik, E., Johnson, W.E., Zhang, Y.P., Ryder, O.A. and O'Brien, S.J. (2001). Molecular phylogenetics and the origins of placental mammals. *Nature* **409**, 614-615.

Napoli, J.L. (1996). Retinoic acid biosynthesis and metabolism. *Federation of American Societies for Experimental Biology Journal* **10**, 993-1001.

Neuweiler, G. (2000). The biology of bats. New York, U.S.A.: Oxford University Press.

Niederreither, K., Vermot, J., Schuhbaur, B., Chambon, P. and Dollé, P. (2002). Embryonic retinoic acid synthesis is required for forelimb growth and anteroposterior patterning in the mouse. *Development* **129**, 3563-3574.

Nolte, M.J., Hockman, D., Cretekos, C.J., Behringer, R.R. and Rasweiler, J.J., IV. (2009). Embryonic staging system for the black mastiff bat, *Molossus rufus* (Molossidae), correlated with structure-function relationships in the adult. *The Anatomical Record* **292**, 155-168.

Norberg, U.M. (1972). Bat wing structures important for aerodynamics and rigidity (Mammalia, Chiroptera). *Zeitschrift für Morphologie und Anthropologie* **73**, 45-61.

Norberg, U.M. and Rayner, J.M.V. (1987). Ecological morphology and flight in bats (Mammalia; Chiroptera): Wing adaptations, flight performance, foraging strategy and echolocation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **316**, 335-427.

Noro, B., Culi, J., McKay, D.J., Zhang, W. and Mann, R.S. (2006). Distinct functions of homeodomain-containing and homeodomain-less isoforms encoded by *homothorax*. *Genes & Development* **20**, 1636-1650.

Norton, P.M. and van der Merwe, M. (1978). Winter activity of bats in a Transvaal highveld cave. *South African Journal of Science* **74**, 216-220.

Nowak, R.M. (1994). Walker's Bats of the World. Baltimore: John Hopkins University Press.

Nowak, R.M. (1999). Walker's mammals of the world. Baltimore: John Hopkins University Press.

Nudds, R. (2007). The wingtip fold of the bat *Miniopterus schreibersii:* A novel mechanism for thrust generation during slow-flight? *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology* **146**, S115.

Operon. (2007a). OpArray™ Protocol. *Operon.*

Operon. *OpArray™ Resource File Downloads* [Online]. Available: https://www.operon.com/arrays/OpArray_downloads.php. June 2007

Oulad-Abdelghani, M., Chazaud, C., Bouillet, P., Sapin, V., Chambon, P. and Dollé, P. (1997). *Meis2*, a novel mouse *Pbx*-related homeobox gene induced by retinoic acid during differentiation of P19 embryonal carcinoma cells. *developmental Dynamics* **210**, 173-183.

Pakrasi, P.L. and Tiwari, A. (2007). Preimplantation embryonic development in a tropical vespertillionid bat. *Journal of Zoology* **272**, 415-422.

Pinheiro, J., Bates, D.M., DebRoy, S., Sarkar, D. and The R Core team. (2007). *nlme*: Linear and nonlinear mixed effects models. R package. **Ver. 3.1-85.**

QIAGEN®. (2006). RNA*later*® Handbook. *QIAGEN®*.

Qin, P., Cimildoro, R., Kochhar, D.M., Soprano, K.J. and Soprano, D.R. (2002). PBX, MEIS, and IGF-I are potential mediators of retinoic acid-induced proximodistal limb reduction defects. *Teratology* **66**, 224-234.

R Development Core Team. (2008). *R: A language and environment for statistical computing*: R Foundation for Statistical Computing. **Ver. 2.6.0.**

Racey, P.A. (1969). Diagnosis of pregnancy and experimental extension of gestation in the pipistrelle Bat, *Pipistrellus Pipistrellus*. *Journal of Reproduction and Fertility* **19**, 465-474.

Racey, P.A. (1973). Environmental factors affecting the length of gestation in heterothermic bats. *Journal of Reproduction and Fertility. Supplement* **19**, 175-189.

Racey, P.A. (1979). The prolonged storage and survival of spermatozoa in Chiroptera. *Journal of Reproduction and Fertility* **56**, 391-402.

Rasweiler IV, J.J. (1979). Early embryonic development and implantation in bats. *Journal of Reproduction and Fertility* **56**, 403-416.

Rasweiler, J.J., IV and Badwaik, N.K. (1997). Delayed development in the short-tailed fruit bat, *Carollia perspicillata*. *Journal of Reproduction and Fertility* **109**, 7-20.

Ray, R. and Capecchi, M.R. (2008). An examination of the Chiropteran HoxD locus from an evolutionary perspective. *Evolution & Development* **10**, 657-670.

Reiter, G. (2004). Postnatal growth and reproductive biology of *Rhinolophus hipposideros* (Chiroptera: Rhinolophidae). *Journal of Zoology* **262**, 231-241.

Renn, S.C.P., Aubin-Horth, N. and Hofmann, H.A. (2004). Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. *BMC Genomics* **5**.

Reno, P.L., McCollum, M.A., Cohn, M.J., Meindl, R.S., Hamrick, M.W. and Lovejoy, C.O. (2008). Patterns of correlation and covariation of anthropoid distal forelimb segments correspond to *Hoxd* expression territories. *Journal of Experimental Zoology Part B: Molecular Development and Evolution* **310B**, 240-258.

Richardson, E.G. (1977). The biology and evolution of the reproductive cycle of *Miniopterus schreibersii* and *M. australis* (Chiroptera: *verspertilionidae*). *Journal of Zoology: Proceedings of the Zoological Society of London* **183**, 353-375.

Rifkin, D.B. (2006). Latent transforming growth factor-β (TGF-β) binding proteins: orchestrators of TGF-β availability. *Journal of Biological Chemistry* **280**, 7409-7412.

Ritchie, M.E., Silver, J., Oshlack, A., Holmes, M., Diyagama, D., Holloway, A. and Smyth, G.K. (2007). A comparison of background correction methods for two-colour microarrays. *Bioinformatics* **23**, 2700-2707.

Rodrigues, L. and Palmeirim, J.M. (2007). Migratory behaviour of the Schreiber's bat: when, where and why do cave bats migrate in a Mediterranean region? *Journal of Zoology* **OnlineEarly Articles**.

Salsi, V., Vigano, M.A., Cocchiarella, F., Mantovani, R. and Zappavigna, V. (2008). Hoxd13 binds *in vivo* and regulates the expression of genes acting in key pathways for early limb and skeletal patterning. *Developmental Biology* **317**, 497-507.

Sears, K.E. (2008). Molecular determinants of bat wing development. *Cells Tissues Organs* **187**, 6-12.

Sears, K.E., Behringer, R.R., Rasweiler, J.J., IV and Niswander, L. (2006). Development of bat flight: Morphologic and molecular evolution of bat wing digits. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 6581-6586.

Sehgal, M.S.B., Gondal, I. and Dooley, L.S. (2005). Collateral missing value imputation: a new robust missing value

estimation algorithm for microarray data. *Bioinformatics* **21**, 2417-2423.

Shalon, D., Smith, S.J. and Brown, P.O. (1996). A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Research* **6**, 639-645.

Shanmugam, K., Green, N.C., Rambaldi, I., Saragovi, H.U. and Featherstone, M.S. (1999). PBX and MEIS as Non-DNA-Binding Partners in Trimeric Complexes with HOX Proteins. *Molecular and Cellular Biology* **19**, 7577-7588.

Shibata, S., Fukada, K., Imai, H., Abe, T. and Yamashita, Y. (2003). *In situ* hybridization and immunohistochemistry of versican, aggrecan and link protein, and histochemistry of hyaluronan in the developing mouse limb bud cartilage. *Journal of Anatomy* **203**, 425-432.

Shou, S., Scott, V., Reed, C., Hitzemann, R. and Stadler, H.S. (2005). Transcriptome analysis of the murine forelimb and hindlimb autopod. *developmental Dynamics* **234**, 74-89.

Shubin, N., Tabin, C.J. and Carroll, S. (1997). Fossils, genes and the evolution of animal limbs. *Nature* **388**, 639-648.

Simmons, N.B. (2005). Order Chiroptera. In *Mammal species of the world: A taxonomic and geographic reference*, vol. 1 (ed. D.E. Wilson and D.M. Reeder), pp. 312-529. Baltimore: John Hopkins University Press.

Simmons, N.B. and Geisler, J.H. (1998). Phylogenetic relationships of *Icaronycteris*, *Archaeonycteris*, *Hassianycteris*, and *Palaeochiropteryx* to extant bat lineages, with comments on the evolution of echolocation and foraging strategies in Microchiroptera. *Bulletin of the American Museum of Natural History* **235**, 1-182.

Simmons, N.B., Seymour, K.L., Habersetzer, J. and Gunnell, G.F. (2008). Primitive Early Eocene bat from Wyoming and the evolution of flight and echolocation. *Nature* **451**, 818-821.

Simpson, G.G. (1953). The major features of evolution. New York: Columbia University Press.

Skinner, J.D. and Smithers, R.H.N. (1990). The mammals of the Southern African subregion. Pretoria: University of Pretoria.

Smith, C.M., Finger, J.H., Hayamizu, T.F., McCright, I.J., Eppig, J.T., Kadin, J.A., Richardson, J.E. and Ringwald, M. (2007). The mouse Gene Expression Database (GXD): 2007 update. *Nucleic Acids Research* **35**, D618-D623.

Smyth, G.K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3**, Article 3.

Smyth, G.K., Michaud, J. and Scott, H.S. (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**, 2067-2075.

Smyth, G.K. and Speed, T.P. (2003). Normalization of cDNA microarray data. *Methods* **31**, 265-273.

Smyth, G.K., Wettenhall, J.M., Yang, Y.H. and Morgan, M. (2004). *convert*: Converting between microarray data classes. R package. **Ver. 1.11.0.**

Speakman, J.R. (2001). The evolution of flight and echolocation in bats: another leap in the dark. *Mammal Review* **31**, 111-130.

Spillmann, F.R. (1925). Beiträge zur kenntnis des fluges der Fledermäuse und der Ontogenetischen entwicklung ihrer Flugapparate. *Acta Zoologica*, 217-223.

Spitz, F., Gonzalez, F. and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* **113**, 405-417.

StatSoft Inc. (2007). *STATISTICA*: Data analysis software system. **Ver. 8.0.**

Steele-Perkins, G., Plachez, C., Butz, K.G., Yang, G., Bachurski, C.J., Kinsman, S.L., Litwack, E.D., Richards, L.J. and Gronostajski, R.M. (2005). The transcription factor gene *Nfib* is essential for both lung maturation and brain development. *Molecular and Cellular Biology* **25**, 685-698.

Sterner-Kock, A., Thorey, I.S., Koli, K., Wempe, F., Otte, J., Bangsow, T., Kuhlmeier, K., Kirchner, T., Jin, S., Keski-Oja, J. et al. (2002). Disruption of the gene encoding the latent transforming growth factor-beta binding protein 4 (LTBP-4) causes abnormal lung development,factor cardiomyopathy, and colorectal cancer. *Genes & Development* **16**, 2264-2273.

Stockwell, E.F. (2001). Morphology and flight manoeuverability in New World leaf-nosed bats (Chiroptera: Phyllostomidae). *Journal of Zoology* **254**, 505-514.

Stoffberg, S., Miller-Butterworth, C.M. and Jacobs, D.S. (2004). Field identification of two morphologically similar bats, *Miniopterus schreibersii natalensis* and *Miniopterus fraterculus* (Chiroptera: Verspertilionidae). *African Zoology* **39**, 47-53.

Studier, E.H., Procter, J.W. and Howell, D.J. (1970). Diurnal body weight loss and tolerance of weight loss in five species of *Myotis*. *Journal of Mammalogy* **51**, 302-309.

Suzuki, T., Hasso, S.M. and Fallon, J.F. (2008). Unique SMAD1/5/8 activity at the phalanx-forming region determines digit identity. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 4185-4190.

Swartz, S.M. (1999). Allometric patterning in the limb skeleton of bats: Implications for the mechanics and energetics of powered flight. *Journal of Morphology* **234**, 277-294.

Swartz, S.M., Bishop, K. and Aguirre, M.-F.I. (2006). Dynamic complexity of wing form in bats: Implications for flight performance. In *Functional and evolutionary ecology of bats*, (ed. A. Zubaid G.F. McCracken and T.H. Kunz), pp. 110-130. Oxford: Oxford University Press.

Tabin, C.J., Carroll, S.B. and Panganiban, G. (1999). Out on a limb: Parallels in vertebrate and invertebrate limb patterning and the origin of appendages. *American Zoologist* **39**, 650-663.

Tarchini, B. and Duboule, D. (2006). Control of *Hoxd* genes' collinearity during early limb development. *Developmental Cell* **10**, 93-103.

Tarchini, B., Duboule, D. and Kmita, M. (2006). Regulatory constraints in the evolution of the tetrapod limb anterior-posterior polarity. *Nature* **443**, 985-988.

Taylor, P.J. (2000). Bats of Southern Africa: Guide to biology, identification and conservation. Scottsville: University of Natal Press.

Teeling, E.C., Scally, M., Kao, D.J., Romagnoli, M.L., Springer, M.S. and Stanhope, M.J. (2000). Molecular evidence regarding the origin of echolocation and fight in bats. *Nature* **403**, 188-192.

ten Berge, D., Brouwer, A., Korving, J., Martin, J.F. and Meijlink, F. (1998). Prx1 and Prx2 in skeletogenesis: roles in the craniofacial region, inner ear and limbs. *Development* **125**, 3831-3842.

Tokita, M. (2006). Normal embryonic development of the Japanese pipistrelle, *Pipistrellus abramus*. *Zoology* **109**, 137-147.

Troyanskaya, O.G., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D. and Altman, R.B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics* **17**, 520-525.

Valentine, J.W., Jablonski, D. and Erwin, D.H. (1999). Fossils, molecules and embryos: new perspectives on the Cambrian explosion. *Development* **126**, 851-859.

van Aarde, R.J., van der Merwe, M. and Skinner, D.C. (1994). Progesterone concentrations and contents in the plasma, ovary, adrenal gland and placenta of the pregnant Natal clinging bat *Miniopterus schreibersii natalensis*. *Journal of Zoology: Proceedings of the Zoological Society of London* **232**, 457-464.

van der Merwe, M. (1973a). Aspects of social behaviour of the Natal clinging bat Miniopterus schreibersii natalensis (A. Smith. 1934). *Mammalia* **37**, 379-389.

van der Merwe, M. (1973b). Aspects of temperature and humidity in preferred hiberanation sites of the Natal clinging bat *Miniopterus schreibersi natalensis* (A. Smith, 1934). *Zoologica Africana* **8**, 12.

van der Merwe, M. (1973c). Aspects of the hibernation and winter activity of the Natal clinging bat, *Miniopterus schreibersi natalensis* (A. Smith 1834), on the Transvaal highveld. *South African Journal of Science* **69**, 116-118.

van der Merwe, M. (1975). Preliminary study of the annual movements of the Natal clinging bat. *South African Journal of Science* **71**, 237-241.

van der Merwe, M. (1979). Foetal growth curves and seasonal breeding in the Natal clinging bat *Miniopterus schreibersii natalensis*. *South African Journal of Zoology* **14**, 17-21.

van der Merwe, M. (1986). Reproductive strategy of *Miniopterus schreibersii natalensis*. *Cimbebasia. Series A, Natural History* **8**, 107-111.

van der Merwe, M. (1996). Amniogenesis in Schreiber's long-fingered bat *Miniopterus schreibersii natalensis*. *South African Journal of Zoology* **14**, 17-21.

Visel, A., Thaller, C. and Eichele, G. (2004). GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Research* **32**, D552-D556.

Wallace, G.I. (1978). A histological study of the early stages of pregnancy in the bent-winged bat (*Miniopterus schreibersii*) in north-eastern New South Wales, Australia (30° 27' S). *Journal of Zoology: Proceedings of the Zoological Society of London* **185**, 519-537.

Wanek, N., Muneoka, K., Holler-Dinsmore, G., Burton, R. and Bryant, S.V. (1989). A staging system for mouse limb development. *Journal of Experimental Zoology* **249**, 41-49.

Watanabe, H. and Yamada, Y. (1999). Mice lacking link protein develop dwarfism and craniofacial abnormalities. *Nature Genetics* **21**, 150-151.

Weatherbee, S.D., Behringer, R.R., Rasweiler, J.J., IV and Niswander, L. (2006). Interdigital webbing retention in bat wings illustrates genetic changes underlying amniote limb diversification. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 15103-15107.

Wernisch, L. (2003). *YASMA*: Yet another statistical microarray analysis. R package. **Ver. 0.20.**

Wijchers, P.J.E.C. (2005). Forkhead transcription factors in brain development: 'Fox hunting in midbrain dopaminergic neurons'. In *Pharmacology and Anatomy*, vol. PhD. Utrecht, The Netherlands: Universiteit Utrecht.

Xiao, Y. and Yang, J.Y.H. (2007). *DEDS*: Differential expression via distance summary for microarray data. R package. **Ver. 1.12.0.**

Xu, X., Weinstein, M., Li, C., Naski, M., Cohen, R.I., Ornitz, D.M., Leder, P. and Deng, C. (1998). Fibroblast growth factor receptor 2 (FGFR2)-mediated reciprocal regulation loop between FGF8 and FGF10 is essential for limb induction. *Development* **125**, 753-765.

Yang, L., Cai1, C.-L., Lin, L., Qyang, Y., Chung, C., Monteiro, R.M., Mummery, C.L., Fishman, G.I., Cogen, A. and Evans, S. (2006). Isl1Cre reveals a common Bmp pathway in heart and limb development. *Development* **133**, 1575-1585.

Yang, Y., Hwang, C.K., D'Souza, U.M., Lee, S.-H., Junn, E. and Mouradian, M.M. (2000). Three-amino acid Extension Loop Homeodomain Proteins Meis2 and TGIF Differentially Regulate Transcription. *Journal of Biological Chemistry* **275**, 20734-20741.

Yang, Y.H. (2008). *marray*: Exploratory analysis for two-color spotted microarray data. R packages. **Ver. 1.15.1.**

Yang, Y.H. and Speed, T.P. (2002). Design issues for cDNA microarray experiments. *Nature Reviews. Genetics* **3**, 579-588.

Yang, Y.H. and Thorne, N.P. (2003). Normalization for two-color cDNA microarray data. In *Science and Statistics: A Festschrift for Terry Speed Speed*, vol. 40 (ed. D.R. Goldstein), pp. 403-418.

Yang, Y.H., Xiao, Y. and Segal, M.R. (2005). Identifying differentially expressed genes from microarray experiments via statistic synthesis. *Bioinformatics* **21**, 1084-1093.

Yashiro, K., Zhao, X., Uehara, M., Yamashita, K., Nishijima, M., Nishino, J., Saijoh, Y., Sakai, Y. and Hamada, H. (2004). Regulation of retinoic acid distribution is required for proximodistal patterning and outgrowth of the developing mouse limb. *Developmental Cell* **6**, 411-422.

Yi, S.E., Daluiski, A., Pederson, R., Rosen, V. and Lyons, K.M. (2000). The type I BMP receptor BMPRIB is required for chondrogenesis in the mouse limb. *Development* **127**, 621-630.

Yoon, B.S. and Lyons, K.M. (2004). Multiple functions of BMPs in chondrogenesis. *Journal of Cellular Biochemistry* **93**, 93-103.

Yoon, B.S., Ovchinnikov, D.A., Yoshii, I., Mishina, Y., Behringer, R.R. and Lyons, K.M. (2005). *Bmpr1a* and *Bmpr1b* have overlapping functions and are essential for chondrogenesis *in vivo*. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 5062-5067.

Zakany, J. and Duboule, D. (1996). Synpolydactyly in mice with a targeted deficiency in the *HoxD* complex. *Nature* **384**, 69-71.

Zakany, J. and Duboule, D. (2007). The role of *Hox* genes during vertebrate limb development. *Current Opinion in Genetics & Development* **17**, 359-366.

Zakany, J., Fromental-Ramain, C., Warot, X. and Duboule, D. (1997). Regulation of number and size of digits by posterior *Hox* genes: a dose dependent mechanism with potential evolutionary implications. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 13695-13700.

Zakany, J., Kmita, M. and Duboule, D. (2004). A dual role for *Hox* genes in limb anterior-posterior asymmetry. *Science* **304**, 1669-1672.

Zakany, J., Zacchetti, G. and Duboule, D. (2007). Interactions between *Hoxd* and *Gli3* genes control the limb apical ectodermal ridge via *Fgf10*. *Developmental Biology* **306**, 883-893.

Zhang, J., Zhang, S. and Jones, G. *University of Bristol Biological Sciences' Guide to Chinese Bats* [Online]. Available: http://www.bio.bris.ac.uk/research/bats/China%20bats/pipistrellusabramus.htm.

Zhang, Z., Schwartz, S., Wagner, L. and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* **7**, 203-214.

Appendix A

R Script

R SCRIPT INDEX


```
#=============================================================================
#                                   R SCRIPT
#=============================================================================
# WRITTEN BY:         Mandy Mason
# DATE:               23 February 2008
# REVISION DATE:      31 July 2008
#
# LANGUAGE:           R version 2.6.0
# PACKAGES NEEDED:    - limma
#                     - marray
#                     - convert
#                     - stats
#                     - agce
#                     - impute
#
# SOURCE DATA TYPE:   Genepix mouse/bat .gal files
# DESCIRPTION:        This script contains functions for normalisation,
#                     preprocessing, batch correction and differential testing
#                     between groups.
#=============================================================================
# OPEN LIBRARIES
#-----------------------------------------------------------------------------

library(limma)
library(marray)
library(convert)
library(stats)
library(agce)
library(impute)
library(annotationTools)
library(yasma)
library(dprep)
library(DEDS)

#=============================================================================
# SET WORKING DIRECTORY
#-----------------------------------------------------------------------------
setwd("C:\\Program Files\\R\\Ultimate Analysis\\R Data Dir")
```

```
################################################################################
# STEP 1:    READ IN NEW DATASET                                               #
################################################################################
targets <- readTargets("Targets.txt")
RG      <- read.maimages(targets$FileName,source="genepix.median",
                         wt.fun=wtflags(weight=0, cutoff=-75))
RG$genes      <- readGAL("FinalGAL.gal")
RG$printer    <- getLayout(RG$genes)
spottypes     <- readSpotTypes("Spot_Types.txt")
RG$genes$Status     <- controlStatus(spottypes,RG)
RG$genes$weights    <- RG$weights

################################################################################
# STEP 2:    VISULISE AND REMOVE DYE MARKERS                                   #
################################################################################
# EXTRACT DYE MARKERS AND CREATE A NEW RGList(RGdm) AND MAList (MAdm)
#-------------------------------------------------------------------------------
gDye <- subset( RG$G, RG$genes$Status=="Dye",
                select = c( "16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                            "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                            "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                            "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))


rDye <- subset( RG$R, RG$genes$Status=="Dye",
                select = c( "16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                            "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                            "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                            "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))

RGdm   <- new("RGList")
RGdm$R <- rDye
RGdm$G <- gDye

#================================================================================
# VISULISE DYE MARKER SIGNAL
#--------------------------------------------------------------------------------
par(mfrow=c(1,1))
mlim <- c(-7,7)
alim <- c(0, 20)

boxplot(log2(RGdm$R[,1]),log2(RGdm$R[,5]),log2(RGdm$R[,9]),log2(RGdm$R[,13]),
        log2(RGdm$R[,2]),log2(RGdm$R[,6]),log2(RGdm$R[,10]),log2(RGdm$R[,14]),
        log2(RGdm$R[,3]),log2(RGdm$R[,7]),log2(RGdm$R[,11]),log2(RGdm$R[,15]),
        log2(RGdm$R[,4]),log2(RGdm$R[,8]),log2(RGdm$R[,12]),log2(RGdm$R[,16]),
        ylim = alim, col= "red")

boxplot(log2(RGdm$G[,1]),log2(RGdm$G[,5]),log2(RGdm$G[,9]),log2(RGdm$G[,13]),
        log2(RGdm$G[,2]),log2(RGdm$G[,6]),log2(RGdm$G[,10]),log2(RGdm$G[,14]),
        log2(RGdm$G[,3]),log2(RGdm$G[,7]),log2(RGdm$G[,11]),log2(RGdm$G[,15]),
        log2(RGdm$G[,4]),log2(RGdm$G[,8]),log2(RGdm$G[,12]),log2(RGdm$G[,16]),
        ylim = alim, col= "green", add=TRUE)

boxplot(MAdm$A[,1],MAdm$A[,5],MAdm$A[,9],MAdm$A[,13],MAdm$A[,2],MAdm$A[,6],
        MAdm$A[,10],MAdm$A[,14],MAdm$A[,3],MAdm$A[,7],MAdm$A[,11],MAdm$A[,15],
        MAdm$A[,4],MAdm$A[,8],MAdm$A[,12],MAdm$A[,16], ylim = alim, col= "blue")

boxplot(MAdm$M[,1],MAdm$M[,5],MAdm$M[,9],MAdm$M[,13],MAdm$M[,2],MAdm$M[,6],
        MAdm$M[,10],MAdm$M[,14],MAdm$M[,3],MAdm$M[,7],MAdm$M[,11],MAdm$M[,15],
        MAdm$M[,4],MAdm$M[,8],MAdm$M[,12],MAdm$M[,16], ylim = mlim, col= "yellow")
```

```
#=============================================================================
# DYE MARKERS STATISTICS
#-----------------------------------------------------------------------------
dmR <- rbind(log2(RGdm$R[,1]),log2(RGdm$R[,2]),log2(RGdm$R[,3]),log2(RGdm$R[,4]),
             log2(RGdm$R[,5]),log2(RGdm$R[,6]),log2(RGdm$R[,7]),log2(RGdm$R[,8]),
             log2(RGdm$R[,9]),log2(RGdm$R[,10]),log2(RGdm$R[,11]),log2(RGdm$R[,12]),
             log2(RGdm$R[,13]),log2(RGdm$R[,14]),log2(RGdm$R[,15]),log2(RGdm$R[,16]))
dmR <- as.vector(dmR)
summary(dmR)

dmG <- rbind(log2(RGdm$G[,1]),log2(RGdm$G[,2]),log2(RGdm$G[,3]),log2(RGdm$G[,4]),
             log2(RGdm$G[,5]),log2(RGdm$G[,6]),log2(RGdm$G[,7]),log2(RGdm$G[,8]),
             log2(RGdm$G[,9]),log2(RGdm$G[,10]),log2(RGdm$G[,11]),log2(RGdm$G[,12]),
             log2(RGdm$G[,13]),log2(RGdm$G[,14]),log2(RGdm$G[,15]),log2(RGdm$G[,16]))
dmG <- as.vector(dmG)
summary(dmG)

dmA <- rbind(MAdm$A[,1],MAdm$A[,2],MAdm$A[,3],MAdm$A[,4],MAdm$A[,5],MAdm$A[,6],
             MAdm$A[,7],MAdm$A[,8],MAdm$A[,9],MAdm$A[,10],MAdm$A[,11],MAdm$A[,12],
             MAdm$A[,13],MAdm$A[,14],MAdm$A[,15],MAdm$A[,16])
dmA <- as.vector(dmA)
summary(dmA)

dmM <- rbind(MAdm$M[,1],MAdm$M[,2],MAdm$M[,3],MAdm$M[,4],MAdm$M[,5],MAdm$M[,6],
             MAdm$M[,7],MAdm$M[,8],MAdm$M[,9],MAdm$M[,10],MAdm$M[,11],MAdm$M[,12],
             MAdm$M[,13],MAdm$M[,14],MAdm$M[,15],MAdm$M[,16])
dmM <- as.vector(dmM)
summary(dmM)

mean(dmR)
sd(dmR)
mean(dmG)
sd(dmG)
mean(dmA)
sd(dmA)
mean(dmM)
sd(dmM)

#=============================================================================
# REMOVE DYE MARKERS
#       - Replace with NA
#-----------------------------------------------------------------------------
RGd <- RG

row <- dim(RGd)[1]
col <- dim(RGd)[2]

x=1
y=1

while(y<=col)
{
    x=1
    while (x<=row)
    {
        if  (RGd$genes[x,6]=="Dye")
        {
            RGd$R[x,y]  = NA
            RGd$G[x,y]  = NA
            RGd$Rb[x,y] = NA
            RGd$Gb[x,y] = NA
        }
    x=x+1
    }
    y=y+1
}
source("remove.NAAdapt.R")
RGdm <- RG.MA(remove.NA(MA.RG(RGd)))
```

```
#############################################################################
# STEP 3:       NORMALISATION                                               #
#               - Norm exp background correction                            #
#               - Robust-spline within arrays                               #
#               - R-Quantile between arrays                                 #
#############################################################################
source("RobustSpline.R")

RGbc   <- backgroundCorrect(RGd, method="normexp", offset = 32)
MAnw   <- normalizeWithinArrays(RGbc, method="robust", maxit=52, weights=TRUE)
MAnorm <- normalizeBetweenArrays(MAnw, method="Rquantile")

#============================================================================
# COERCE DATA INTO DIFFERENT CLASSES
#----------------------------------------------------------------------------
MA     <- MA.RG(RG, bc.method="none")
MAd    <- MA.RG(RGd, bc.method="none")
MAbc   <- MA.RG(RGbc, bc.method="none")

RGbc   <- RG.MA(MAbc)
RGnw   <- RG.MA(MAnw)
RGnorm <- RG.MA(MAnorm)

#============================================================================
# VISULIZE DATA BEFORE AND AFTER NORMALISATION
#      - MA plots
#      - Density plot
#      - Boxplot of M values
#      - Boxplot of A values
#----------------------------------------------------------------------------
maMA     <- as(MA, "marrayNorm")
maMAnorm <- as(MAnorm, "marrayNorm")

par(mfrow=c(1,1))
mlim = c(-7,7)
alim = c(4, 17)

par(mfrow=c(1,1))
#----------------------------------------------------------------------------
# Set x = to array number
#----------------------------------------------------------------------------
x = 1

maPlot(maMA[,x], lines.func=NULL, key.func=NULL, lwd=1, ylim=mlim)
points(maMA[,x], subset=abs(maMA@maM[,x])>1, col="orange", pch=20)
points(maMA[,x], subset=abs(maMA@maM[,x])>2, col="red", pch=20)
points(maMA[,x], subset=maMA@maGnames@maInfo[,6]=="Dye", col="white", pch=20)
points(maMA[,x], subset=maMA@maGnames@maInfo[,6]=="Null", col="blue", pch=20)
points(maMA[,x], subset=maMA@maGnames@maInfo[,6]=="Negative", col="light blue",
       pch=20)
points(maMA[,x], subset=maMA@maGnames@maInfo[,6]=="Control", col="yellow", pch=20)
points(maMA[,x], subset=maMA@maGnames@maInfo[,6]=="St60", col="green1", pch=18)
points(maMA[,x], subset=maMA@maGnames@maInfo[,6]=="St70", col="green2", pch=18)
points(maMA[,x], subset=maMA@maGnames@maInfo[,6]=="St80", col="green3", pch=18)
points(maMA[,x], subset=maMA@maGnames@maInfo[,6]=="St90", col="green4", pch=18)
points(maMA[,x], subset=maMA@maGnames@maInfo[,6]=="St100", col="green", pch=18)
savePlot(filename = "maMA(x)", type = c("bmp"),device = dev.cur(),
         restoreConsole = TRUE)

maPlot(maMAnorm[,x], lines.func=NULL, key.func=NULL, lwd=1, ylim=mlim)
points(maMAnorm[,x], subset=abs(maMAnorm@maM[,x])>1, col="orange", pch=20)
points(maMAnorm[,x], subset=abs(maMAnorm@maM[,x])>2, col="red", pch=20)
points(maMAnorm[,x], subset=maMAnorm@maGnames@maInfo[,6]=="Dye", col="white",
       pch=20)
points(maMAnorm[,x], subset=maMAnorm@maGnames@maInfo[,6]=="Null", col="blue",
       pch=20)
points(maMAnorm[,x], subset=maMAnorm@maGnames@maInfo[,6]=="Negative",
       col="light blue", pch=20)
```

```
points(maMAnorm[,x], subset=maMAnorm@maGnames@maInfo[,6]=="Control", col="yellow",
       pch=20)
points(maMAnorm[,x], subset=maMAnorm@maGnames@maInfo[,6]=="St60", col="green1",
       pch=18)
points(maMAnorm[,x], subset=maMAnorm@maGnames@maInfo[,6]=="St70", col="green2",
       pch=18)
points(maMAnorm[,x], subset=maMAnorm@maGnames@maInfo[,6]=="St80", col="green3",
       pch=18)
points(maMAnorm[,x], subset=maMAnorm@maGnames@maInfo[,6]=="St90", col="green4",
       pch=18)
points(maMAnorm[,x], subset=maMAnorm@maGnames@maInfo[,6]=="St100", col="green",
       pch=18)
savePlot(filename = "maMAnorm(x)", type = c("bmp"),device = dev.cur(),
         restoreConsole = TRUE)

plotDensities(MA)
plotDensities(MAd)
plotDensities(MAbc)
plotDensities(MAnw)
plotDensities(MAnorm)
boxplot(MA$M[,1],MA$M[,5],MA$M[,9],MA$M[,13],MA$M[,2],MA$M[,6],MA$M[,10],MA$M[,14],
        MA$M[,3],MA$M[,7],MA$M[,11],MA$M[,15],MA$M[,4],MA$M[,8],MA$M[,12],MA$M[,16],
        ylim=mlim)

boxplot(MAd$M[,1],MAd$M[,5],MAd$M[,9],MAd$M[,13],MAd$M[,2],MAd$M[,6],MAd$M[,10],
        MAd$M[,14],MAd$M[,3],MAd$M[,7],MAd$M[,11],MAd$M[,15],MAd$M[,4],MAd$M[,8],
        MAd$M[,12],MAd$M[,16], ylim=mlim)

boxplot(MAbc$M[,1],MAbc$M[,5],MAbc$M[,9],MAbc$M[,13],MAbc$M[,2],MAbc$M[,6],
        MAbc$M[,10],MAbc$M[,14],MAbc$M[,3],MAbc$M[,7],MAbc$M[,11],MAbc$M[,15],
        MAbc$M[,4],MAbc$M[,8],MAbc$M[,12],MAbc$M[,16], ylim=mlim)

boxplot(MAnw$M[,1],MAnw$M[,5],MAnw$M[,9],MAnw$M[,13],MAnw$M[,2],MAnw$M[,6],
        MAnw$M[,10],MAnw$M[,14],MAnw$M[,3],MAnw$M[,7],MAnw$M[,11],MAnw$M[,15],
        MAnw$M[,4],MAnw$M[,8],MAnw$M[,12],MAnw$M[,16], ylim=mlim)

boxplot(MAnorm$M[,1],MAnorm$M[,5],MAnorm$M[,9],MAnorm$M[,13],MAnorm$M[,2],
        MAnorm$M[,6],MAnorm$M[,10],MAnorm$M[,14],MAnorm$M[,3],MAnorm$M[,7],
        MAnorm$M[,11],MAnorm$M[,15],MAnorm$M[,4],MAnorm$M[,8],MAnorm$M[,12],
        MAnorm$M[,16], ylim=mlim)

boxplot(MA$A[,1],MA$A[,5],MA$A[,9],MA$A[,13],MA$A[,2],MA$A[,6],MA$A[,10],MA$A[,14],
        MA$A[,3],MA$A[,7],MA$A[,11],MA$A[,15],MA$A[,4],MA$A[,8],MA$A[,12],MA$A[,16],
        ylim=alim)

boxplot(MAd$A[,1],MAd$A[,5],MAd$A[,9],MAd$A[,13],MAd$A[,2],MAd$A[,6],MAd$A[,10],
        MAd$A[,14],MAd$A[,3],MAd$A[,7],MAd$A[,11],MAd$A[,15],MAd$A[,4],MAd$A[,8],
        MAd$A[,12],MAd$A[,16], ylim=alim)

boxplot(MAbc$A[,1],MAbc$A[,5],MAbc$A[,9],MAbc$A[,13],MAbc$A[,2],MAbc$A[,6],
        MAbc$A[,10],MAbc$A[,14],MAbc$A[,3],MAbc$A[,7],MAbc$A[,11],MAbc$A[,15],
        MAbc$A[,4],MAbc$A[,8],MAbc$A[,12],MAbc$A[,16], ylim=alim)

boxplot(MAnw$A[,1],MAnw$A[,5],MAnw$A[,9],MAnw$A[,13],MAnw$A[,2],MAnw$A[,6],
        MAnw$A[,10],MAnw$A[,14],MAnw$A[,3],MAnw$A[,7],MAnw$A[,11],MAnw$A[,15],
        MAnw$A[,4],MAnw$A[,8],MAnw$A[,12],MAnw$A[,16], ylim=alim)

boxplot(MAnorm$A[,1],MAnorm$A[,5],MAnorm$A[,9],MAnorm$A[,13],MAnorm$A[,2],
        MAnorm$A[,6],MAnorm$A[,10],MAnorm$A[,14],MAnorm$A[,3],MAnorm$A[,7],
        MAnorm$A[,11],MAnorm$A[,15],MAnorm$A[,4],MAnorm$A[,8],MAnorm$A[,12],
        MAnorm$A[,16], ylim=alim)
```

```
#############################################################################
# STEP 4:       DETERMINE THRESHOLD OF 'DEAD' SIGNAL                        #
#############################################################################
# PLOT CORRELATION (R2) BETWEEN ALL EXPERIMENTS WHEN LOWER QUANTILES REMOVED
#---------------------------------------------------------------------------
RGnormNA <- RG.MA(remove.NA(MA.RG(RGnorm)))

source("yasmaAdapt.R")

rg.rsq.plot(RGd, p.end=0.99)
rg.rsq.plot(RGnorm, p.end=0.99)

R2Low <- c(rg.rsq(RGnormNA, p.low=0.0),rg.rsq(RGnormNA, p.low=0.1),rg.rsq(RGnormNA,
          p.low=0.2),rg.rsq(RGnormNA, p.low=0.3),rg.rsq(RGnormNA, p.low=0.4),
          rg.rsq(RGnormNA, p.low=0.5), rg.rsq(RGnormNA, p.low=0.6),
          rg.rsq(RGnormNA, p.low=0.7), rg.rsq(RGnormNA, p.low=0.8),
          rg.rsq(RGnormNA, p.low=0.9), rg.rsq(RGnormNA, p.low=0.99),
          rg.rsq(RGnormNA, p.low=0.999))
R2High <- c(rg.rsq(RGnormNA, p.high=0.0),rg.rsq(RGnormNA, p.high=0.1),
          rg.rsq(RGnormNA, p.high=0.2),rg.rsq(RGnormNA, p.high=0.3),
          rg.rsq(RGnormNA, p.high=0.4), rg.rsq(RGnormNA, p.high=0.5),
          rg.rsq(RGnormNA, p.high=0.6), rg.rsq(RGnormNA, p.high=0.7),
          rg.rsq(RGnormNA, p.high=0.8), rg.rsq(RGnormNA, p.high=0.9),
          rg.rsq(RGnormNA, p.high=0.99), rg.rsq(RGnormNA, p.high=0.999))
FractionL <- c(0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.99, 0.999)
FractionH <- c(0.999, 0.99, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0)

par(mfrow=c(1,1))

plot(FractionL,R2Low, pch=19, bty="l", ylim=c(0,1))
lines(FractionL,R2Low,type="l", ljoin = "round", bty="l",col="black", lwd=2)
abline(v=0.30, col="red", lty = "longdash", lwd=2.5)

plot(FractionH, R2High, pch=19, bty="l", ylim=c(0,1), col = "blue")
lines(FractionH,R2High,type="l", ljoin = "round", bty="l",col="blue", lwd=2)
abline(v=0.4, col="red", lty = "longdash", lwd=2.5)

#============================================================================
# CREATE DATAFRAMES OF RED AND GREEN VALUES FOR EACH STATUS GROUP
#      - Background, Empty/Null, Buffer, Negative
#      - Raw, background corrected, normalised
#----------------------------------------------------------------------------
rBack <-remove.NA(RGd$Rb)
gBack <-remove.NA(RGd$Gb)

gNull <- subset(RGd$G, RGd$genes$Status=="Null",
                select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                           "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                           "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                           "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))
rNull <- subset(RGd$R, RGd$genes$Status=="Null",
                select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                           "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                           "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                           "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))
gBuff <- subset(RGd$G, RGd$genes$Status=="Buffer",
                select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                           "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                           "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                           "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))
rBuff <- subset(RGd$R, RGd$genes$Status=="Buffer",
                select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                           "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                           "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                           "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))
gNeg  <- subset(RGd$G, RGd$genes$Status=="Negative",
                select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                           "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
```

```
                          "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                          "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))
rNeg  <- subset(RGd$R, RGd$genes$Status=="Negative",
                select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                          "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                          "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                          "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))

gNullbc <- subset(RGbc$G, RGbc$genes$Status=="Null",
                  select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                            "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                            "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                            "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))
rNullbc <- subset(RGbc$R, RGbc$genes$Status=="Null",
                  select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                             "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                             "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                             "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))
gBuffbc <- subset(RGbc$G, RGbc$genes$Status=="Buffer",
                  select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                             "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                             "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                             "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))

rBuffbc <- subset(RGbc$R, RGbc$genes$Status=="Buffer",
                  select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                             "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                             "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                             "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))
gNegbc  <- subset(RGbc$G, RGbc$genes$Status=="Negative",
                  select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                             "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                             "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                             "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))
rNegbc  <- subset(RGbc$R, RGbc$genes$Status=="Negative",
                  select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                             "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                             "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                             "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))

gNulln <- subset(RGnorm$G, RGnorm$genes$Status=="Null",
                  select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                             "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                             "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                             "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))
rNulln <- subset(RGnorm$R, RGnorm$genes$Status=="Null",
                  select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                             "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                             "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                             "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))
gBuffn <- subset(RGnorm$G, RGnorm$genes$Status=="Buffer",
                  select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                             "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                             "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                             "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))
rBuffn <- subset(RGnorm$R, RGnorm$genes$Status=="Buffer",
                  select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                             "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                             "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                             "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))
gNegn  <- subset(RGnorm$G, RGnorm$genes$Status=="Negative",
                  select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                             "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
                             "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                             "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))
rNegn  <- subset(RGnorm$R, RGnorm$genes$Status=="Negative",
                  select = c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
                             "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
```

```
                                "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
                                "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))

#=============================================================================
# CREATE RGList FOR EACH STATUS GROUP
#       - Background, Empty/Null, Buffer, Negative
#-----------------------------------------------------------------------------
RGb <- new("RGList")
RGb$R = rBack
RGb$G = gBack

RGn <- new("RGList")
RGn$R = rNull
RGn$G = gNull

RGbf <- new("RGList")
RGbf$R = rBuff
RGbf$G = gBuff

RGng <- new("RGList")
RGng$R = rNeg
RGng$G = gNeg

RGnbc <- new("RGList")
RGnbc$R = rNullbc
RGnbc$G = gNullbc

RGbfbc <- new("RGList")
RGbfbc$R = rBuffbc
RGbfbc$G = gBuffbc

RGngbc <- new("RGList")
RGngbc$R = rNegbc
RGngbc$G = gNegbc

RGnn <- new("RGList")
RGnn$R = rNulln
RGnn$G = gNulln

RGbfn <- new("RGList")
RGbfn$R = rBuffn
RGbfn$G = gBuffn

RGngn <- new("RGList")
RGngn$R = rNegn
RGngn$G = gNegn

#=============================================================================
# CREATE MARRAYList FOR EACH STATUS GROUP
#       - Background, Empty/Null, Buffer, Negative
#-----------------------------------------------------------------------------
MAb    <- MA.RG(RGb, bc.method="none")
MAn    <- MA.RG(RGn, bc.method="none")
MAbf   <- MA.RG(RGbf, bc.method="none")
MAng   <- MA.RG(RGng, bc.method="none")
MAnbc  <- MA.RG(RGnbc, bc.method="none")
MAbfbc <- MA.RG(RGbfbc, bc.method="none")
MAngbc <- MA.RG(RGngbc, bc.method="none")
MAnn   <- MA.RG(RGnn, bc.method="none")
MAbfn  <- MA.RG(RGbfn, bc.method="none")
MAngn  <- MA.RG(RGngn, bc.method="none")
```

```
#=============================================================================
# VISULISE LOW INTENSITY DATA BY ARRAY
#-----------------------------------------------------------------------------
par(mfrow=c(1,1))

boxplot(MAb$A[,1],MAb$A[,5],MAb$A[,9],MAb$A[,13],MAb$A[,2],MAb$A[,6],MAb$A[,10],
        MAb$A[,14],MAb$A[,3],MAb$A[,7],MAb$A[,11],MAb$A[,15],MAb$A[,4],MAb$A[,8],
        MAb$A[,12],MAb$A[,16], ylim=alim, col= "blue")
boxplot(MAn$A[,1],MAn$A[,5],MAn$A[,9],MAn$A[,13],MAn$A[,2],MAn$A[,6],MAn$A[,10],
        MAn$A[,14],MAn$A[,3],MAn$A[,7],MAn$A[,11],MAn$A[,15],MAn$A[,4],MAn$A[,8],
        MAn$A[,12],MAn$A[,16], ylim=alim, col= "blue")
boxplot(MAbf$A[,1],MAbf$A[,5],MAbf$A[,9],MAbf$A[,13],MAbf$A[,2],MAbf$A[,6],
        MAbf$A[,10],MAbf$A[,14],MAbf$A[,3],MAbf$A[,7],MAbf$A[,11],MAbf$A[,15],
        MAbf$A[,4],MAbf$A[,8],MAbf$A[,12],MAbf$A[,16], ylim=alim, col= "blue")
boxplot(MAng$A[,1],MAng$A[,5],MAng$A[,9],MAng$A[,13],MAng$A[,2],MAng$A[,6],
        MAng$A[,10],MAng$A[,14],MAng$A[,3],MAng$A[,7],MAng$A[,11],MAng$A[,15],
        MAng$A[,4],MAng$A[,8],MAng$A[,12],MAng$A[,16], ylim=alim, col= "blue")
boxplot(MAnbc$A[,1],MAnbc$A[,5],MAnbc$A[,9],MAnbc$A[,13],MAnbc$A[,2],MAnbc$A[,6],
        MAnbc$A[,10],MAnbc$A[,14],MAnbc$A[,3],MAnbc$A[,7],MAnbc$A[,11],MAnbc$A[,15],
        MAnbc$A[,4],MAnbc$A[,8],MAnbc$A[,12],MAnbc$A[,16], ylim=alim, col= "blue")
boxplot(MAbfbc$A[,1],MAbfbc$A[,5],MAbfbc$A[,9],MAbfbc$A[,13],MAbfbc$A[,2],
        MAbfbc$A[,6],MAbfbc$A[,10],MAbfbc$A[,14],MAbfbc$A[,3],MAbfbc$A[,7],
        MAbfbc$A[,11],MAbfbc$A[,15],MAbfbc$A[,4],MAbfbc$A[,8],MAbfbc$A[,12],
        MAbfbc$A[,16], ylim=alim, col= "blue")
boxplot(MAngbc$A[,1],MAngbc$A[,5],MAngbc$A[,9],MAngbc$A[,13],MAngbc$A[,2],
        MAngbc$A[,6],MAngbc$A[,10],MAngbc$A[,14],MAngbc$A[,3],MAngbc$A[,7],
        MAngbc$A[,11],MAngbc$A[,15],MAngbc$A[,4],MAngbc$A[,8],MAngbc$A[,12],
        MAngbc$A[,16], ylim=alim, col= "blue")
boxplot(MAnn$A[,1],MAnn$A[,5],MAnn$A[,9],MAnn$A[,13],MAnn$A[,2],MAnn$A[,6],
        MAnn$A[,10],MAnn$A[,14],MAnn$A[,3],MAnn$A[,7],MAnn$A[,11],MAnn$A[,15],
        MAnn$A[,4],MAnn$A[,8],MAnn$A[,12],MAnn$A[,16], ylim=alim, col= "blue")
boxplot(MAbfn$A[,1],MAbfn$A[,5],MAbfn$A[,9],MAbfn$A[,13],MAbfn$A[,2],MAbfn$A[,6],
        MAbfn$A[,10],MAbfn$A[,14],MAbfn$A[,3],MAbfn$A[,7],MAbfn$A[,11],MAbfn$A[,15],
        MAbfn$A[,4],MAbfn$A[,8],MAbfn$A[,12],MAbfn$A[,16], ylim=alim, col= "blue")
boxplot(MAngn$A[,1],MAngn$A[,5],MAngn$A[,9],MAngn$A[,13],MAngn$A[,2],MAngn$A[,6],
        MAngn$A[,10],MAngn$A[,14],MAngn$A[,3],MAngn$A[,7],MAngn$A[,11],MAngn$A[,15],
        MAngn$A[,4],MAngn$A[,8],MAngn$A[,12],MAngn$A[,16], ylim=alim, col= "blue")

plotDensities(MAb)

plotDensities(MAn)
plotDensities(MAbf)
plotDensities(MAng)

plotDensities(MAnbc)
plotDensities(MAbfbc)
plotDensities(MAngbc)

plotDensities(MAnn)
plotDensities(MAbfn)
plotDensities(MAngn)

#=============================================================================
# CALCULATE DISTRIBUTION SUMMARIES FOR EACH ARRAY/STATUS GROUP
#-----------------------------------------------------------------------------

bA <- rbind(MAb$A[,1],MAb$A[,2],MAb$A[,3],MAb$A[,4],MAb$A[,5],MAb$A[,6], MAb$A[,7],
            MAb$A[,8],MAb$A[,9],MAb$A[,10],MAb$A[,11],MAb$A[,12],MAb$A[,13],
            MAb$A[,14],MAb$A[,15],MAb$A[,16])
nA <- rbind(MAn$A[,1],MAn$A[,2],MAn$A[,3],MAn$A[,4],MAn$A[,5],MAn$A[,6],MAn$A[,7],
            MAn$A[,8],MAn$A[,9],MAn$A[,10],MAn$A[,11],MAn$A[,12],MAn$A[,13],
            MAn$A[,14],MAn$A[,15],MAn$A[,16])
bfA <- rbind(MAbf$A[,1],MAbf$A[,2],MAbf$A[,3],MAbf$A[,4],MAbf$A[,5],MAbf$A[,6],
            MAbf$A[,7],MAbf$A[,8],MAbf$A[,9],MAbf$A[,10],MAbf$A[,11],MAbf$A[,12],
            MAbf$A[,13],MAbf$A[,14],MAbf$A[,15],MAbf$A[,16])
ngA <- rbind(MAng$A[,1],MAng$A[,2],MAng$A[,3],MAng$A[,4],MAng$A[,5],MAng$A[,6],
            MAng$A[,7],MAng$A[,8],MAng$A[,9],MAng$A[,10],MAng$A[,11],MAng$A[,12],
            MAng$A[,13],MAng$A[,14],MAng$A[,15],MAng$A[,16])
```

```
nbcA  <- rbind(MAnbc$A[,1],MAnbc$A[,2],MAnbc$A[,3],MAnbc$A[,4],MAnbc$A[,5],
            MAnbc$A[,6],MAnbc$A[,7],MAnbc$A[,8],MAnbc$A[,9],MAnbc$A[,10],
            MAnbc$A[,11],MAnbc$A[,12],MAnbc$A[,13],MAnbc$A[,14],MAnbc$A[,15],
            MAnbc$A[,16])
bfbcA <- rbind(MAbfbc$A[,1],MAbfbc$A[,2],MAbfbc$A[,3],MAbfbc$A[,4],MAbfbc$A[,5],
            MAbfbc$A[,6],MAbfbc$A[,7],MAbfbc$A[,8],MAbfbc$A[,9],MAbfbc$A[,10],
            MAbfbc$A[,11],MAbfbc$A[,12],MAbfbc$A[,13],MAbfbc$A[,14],MAbfbc$A[,15],
            MAbfbc$A[,16])
ngbcA <- rbind(MAngbc$A[,1],MAngbc$A[,2],MAngbc$A[,3],MAngbc$A[,4],MAngbc$A[,5],
            MAngbc$A[,6],MAngbc$A[,7],MAngbc$A[,8],MAngbc$A[,9],MAngbc$A[,10],
            MAngbc$A[,11],MAngbc$A[,12],MAngbc$A[,13],MAngbc$A[,14],MAngbc$A[,15],
            MAngbc$A[,16])
nnA <- rbind(MAnn$A[,1],MAnn$A[,2],MAnn$A[,3],MAnn$A[,4],MAnn$A[,5],MAnn$A[,6],
            MAnn$A[,7],MAnn$A[,8],MAnn$A[,9],MAnn$A[,10],MAnn$A[,11],MAnn$A[,12],
            MAnn$A[,13],MAnn$A[,14],MAnn$A[,15],MAnn$A[,16])
bfnA <- rbind(MAbfn$A[,1],MAbfn$A[,2],MAbfn$A[,3],MAbfn$A[,4],MAbfn$A[,5],
            MAbfn$A[,6],MAbfn$A[,7],MAbfn$A[,8],MAbfn$A[,9],MAbfn$A[,10],
            MAbfn$A[,11],MAbfn$A[,12],MAbfn$A[,13],MAbfn$A[,14],MAbfn$A[,15],
            MAbfn$A[,16])
ngnA <- rbind(MAngn$A[,1],MAngn$A[,2],MAngn$A[,3],MAngn$A[,4],MAngn$A[,5],
            MAngn$A[,6],MAngn$A[,7],MAngn$A[,8],MAngn$A[,9],MAngn$A[,10],
            MAngn$A[,11],MAngn$A[,12],MAngn$A[,13],MAngn$A[,14],MAngn$A[,15],
            MAngn$A[,16])

bA    <- as.vector(bA)
nA    <- as.vector(nA)
bfA   <- as.vector(bfA)
ngA   <- as.vector(ngA)
nbcA  <- as.vector(nbcA)
bfbcA <- as.vector(bfbcA)
ngbcA <- as.vector(ngbcA)
nnA   <- as.vector(nnA)
bfnA  <- as.vector(bfnA)
ngnA  <- as.vector(ngnA)


summary(bA)
summary(nA)
summary(bfA)
summary(ngA)
summary(nbcA)
summary(bfbcA)
summary(ngbcA)
summary(nnA)
summary(bfnA)
summary(ngnA)

bAstats    <- boxplot.stats(bA)
nAstats    <- boxplot.stats(nA)
bfAstats   <- boxplot.stats(bfA)
ngAstats   <- boxplot.stats(ngA)
nbcAstats  <- boxplot.stats(nbcA)
bfbcAstats <- boxplot.stats(bfbcA)
ngbcAstats <- boxplot.stats(ngbcA)
nnAstats   <- boxplot.stats(nnA)
bfnAstats  <- boxplot.stats(bfnA)
ngnAstats  <- boxplot.stats(ngnA)

bAstats$stats
nAstats$stats
bfAstats$stats
ngAstats$stats
nbcAstats$stats
bfbcAstats$stats
ngbcAstats$stats
nnAstats$stats
bfnAstats$stats
ngnAstats$stats
```

```
mean(bA)
mean(nA)
mean(bfA)
mean(ngA)
mean(nbcA)
mean(bfbcA)
mean(ngbcA)
mean(nnA)
mean(bfnA)
mean(ngnA)

sd(bA)
sd(nA)
sd(bfA)
sd(ngA)
sd(nbcA)
sd(bfbcA)
sd(ngbcA)
sd(nnA)
sd(bfnA)
sd(ngnA)


sdnnM <-sd(t(MAnn$M))
sdbfnM <-sd(t(MAbfn$M))
sdngnM <-sd(t(MAngn$M))

mean(sdnnM)
mean(sdbfnM)
mean(sdngnM)

All  <- rbind((nnAstats$stats[5]),(bfnAstats$stats[5]),(ngnAstats$stats[5]))
mean(All)
sd(All)
AL   <- mean(All)

par(mfrow=c(1,1))
boxplot(bA, nA, nbcA, nnA, bfA, bfbcA, bfnA, ngA, ngbcA, ngnA)
abline(h=AL, col="red", lty="longdash", lwd=3)

par(mfrow=c(1,1))
boxplot(bA, nA,  bfA, ngA,ylim=alim)

par(mfrow=c(1,1))
boxplot(bA, nnA,  bfnA, ngnA, MAnorm$A[,1],MAnorm$A[,5],MAnorm$A[,9],MAnorm$A[,13],
        MAnorm$A[,2],MAnorm$A[,6],MAnorm$A[,10],MAnorm$A[,14],MAnorm$A[,3],
        MAnorm$A[,7],MAnorm$A[,11],MAnorm$A[,15],MAnorm$A[,4],MAnorm$A[,8],
        MAnorm$A[,12],MAnorm$A[,16], ylim=alim)
abline(h=AL, col="red", lty="longdash", lwd=3)
```

```
#############################################################################
# STEP 5a:      FILTER OUT 'DEAD' SIGNAL                                    #
#               - Remove low intensity signal                               #
#               - Genes < L across all arrays                               #
#############################################################################
# REMOVE MISSING VALUES (DYE MARKERS) FROM DATASET
#----------------------------------------------------------------------------

MAnorm <- remove.NA(MAnorm)

#============================================================================
# SET LIMIT (L) FOR FILTER CUTOFF
#       - AL = A value below which signal is considered not present
#----------------------------------------------------------------------------

AL <- mean(All)

#============================================================================
# SET MAXIMUM ROW NUMBER
#       - max = length of class MA$A - A values
#----------------------------------------------------------------------------

max1 <- length(MAnorm$A[,1])

#============================================================================
# CREATE BINARY TRUE/FALSE ARRAY FOR EACH COLUMN/EXPERIMENT
#       - TRUE = 1 if A value is larger than L
#       - FALSE = 0 if A value is smaller than L
#----------------------------------------------------------------------------

g1  <- (MAnorm$A[,1]<=AL)
g2  <- (MAnorm$A[,2]<=AL)
g3  <- (MAnorm$A[,3]<=AL)
g4  <- (MAnorm$A[,4]<=AL)
g5  <- (MAnorm$A[,5]<=AL)
g6  <- (MAnorm$A[,6]<=AL)
g7  <- (MAnorm$A[,7]<=AL)
g8  <- (MAnorm$A[,8]<=AL)
g9  <- (MAnorm$A[,9]<=AL)
g10 <- (MAnorm$A[,10]<=AL)
g11 <- (MAnorm$A[,11]<=AL)
g12 <- (MAnorm$A[,12]<=AL)
g13 <- (MAnorm$A[,13]<=AL)
g14 <- (MAnorm$A[,14]<=AL)
g15 <- (MAnorm$A[,15]<=AL)
g16 <- (MAnorm$A[,16]<=AL)

#============================================================================
# CREATE VECTOR x FOR EACH ROW (GENE)
#       - x = sum of all arrays for each gene
#----------------------------------------------------------------------------

y <- c(mean(g1),mean(g2),mean(g3),mean(g4),mean(g5),mean(g6),mean(g7),mean(g8),
       mean(g9),mean(g10),mean(g11),mean(g12),mean(g13),mean(g14),mean(g15),
       mean(g16))

mean(y)
sd(y)

x <- g1+g2+g3+g4+g5+g6+g7+g8+g9+g10+g11+g12+g13+g14+g15+g16
z <- x
```

```
#=============================================================================
# TEST IF VECTOR x IS LARGER OR EQUAL TO 0 AND CREATE VECTOR z
#       - If x value > 0 then one or more values in array is > L therefore make z
#         value = to 1)
#       - If x value = 0 then all values in array are < L therefore make z
#         value= NA)
#-----------------------------------------------------------------------------

i=1
while(i<=(max1))
{
    if (x[i]<16)
    {
        z[i]=1
    }
    else
    {
        z[i]=NA
    }
    i=i+1
}

#=============================================================================
# CREATE NEW CLASSES: MAlist (MAf) and RGlist (RGf)
#-----------------------------------------------------------------------------

MAf <- MAnorm

#=============================================================================
# MULTIPLY NEW OBJECTS BY BINARY VECTOR z
#       - If z value = 1 object row values remain unchanged
#       - If z value = NA object row values become NA
#-----------------------------------------------------------------------------

MAf$A <- MAnorm$A*z
MAf$M <- MAnorm$M*z

RGf <- RG.MA(remove.NA(MAf))

#############################################################################
# STEP 5b:      FILTER OUT HEAVILY WEIGHTED ROWS                            #
#               - Bad signal across arrays                                  #
#               - Remove genes that are weighted bad in f or more arrays    #
#############################################################################
# DEFINE CUTOFF FOR WEIGHTED VALUES
#       - f = the limit for number of arrays/experiments that are not weighted (1)
#             - filtercutoff
#       - rows with number of good weights > f will be retained
#       - rows with number of good weights < f will be removed
#-----------------------------------------------------------------------------

f <- 12

#=============================================================================
# DETERMINE SUM OF WEIGHTS FOR EACH ROW
#-----------------------------------------------------------------------------

w <- (MAf$weights[,1]+MAf$weights[,2]+MAf$weights[,3]+MAf$weights[,4]
     +MAf$weights[,5]+MAf$weights[,6]+MAf$weights[,7]+MAf$weights[,8]
     +MAf$weights[,9]+MAf$weights[,10]+MAf$weights[,11]+MAf$weights[,12]
     +MAf$weights[,13]+MAf$weights[,14]+MAf$weights[,15]+MAf$weights[,16])

d <- w
```

```
#=============================================================================
# SET MAXIMUM ROW NUMBER
#       - max = length of class MA$A - A values
#-----------------------------------------------------------------------------

max2 <- length(MAf$M[,1])

#=============================================================================
# TEST IF VECTOR weighttest IS LARGER OR SMALLER THAN 12 AND CREATE VECTOR d
#       - If w value > f make n value = 1
#       - If w value < f make n value = NA
#-----------------------------------------------------------------------------

n=1
while(n<=(max2))
{
    if (w[n]<=f)
    {
        d[n]=NA
    }
    else d[n]=1
    n=n+1
}

#=============================================================================
# CREATE NEW CLASSES: MAList (MAw)AND RGList (RGw)
#-----------------------------------------------------------------------------

MAw <- MAf

#=============================================================================
# MULTIPLY NEW MA OBJECTS BY VECTOR d
#       - If d value = 1 object row values remain unchanged
#       - If d value = NA object row values become NA
#-----------------------------------------------------------------------------

MAw$A <- MAf$A*d
MAw$M <- MAf$M*d
MAw$weights <- MAf$weights*d

#############################################################################
# STEP 5c:      VISULISE DATA DISTRIBUTIONS THROUGH ALL FILTERING STEPS      #
#               - Density distribution plots                                 #
#               - Boxplots of M values                                       #
#               - Boxplots of A values                                       #
#############################################################################
# DENSITY DISTRIBUTIONS
#-----------------------------------------------------------------------------

plotDensities(MAnorm)
plotDensities(MAf)
plotDensities(MAw)

#=============================================================================
# BOXPLOTS OF M VALUES
#-----------------------------------------------------------------------------

boxplot(MAnorm$M[,1],MAnorm$M[,5],MAnorm$M[,9],MAnorm$M[,13],MAnorm$M[,2],
        MAnorm$M[,6],MAnorm$M[,10],MAnorm$M[,14],MAnorm$M[,3],MAnorm$M[,7],
        MAnorm$M[,11],MAnorm$M[,15],MAnorm$M[,4],MAnorm$M[,8],MAnorm$M[,12],
        MAnorm$M[,16], ylim=mlim)
boxplot(MAf$M[,1],MAf$M[,5],MAf$M[,9],MAf$M[,13],MAf$M[,2],MAf$M[,6],MAf$M[,10],
        MAf$M[,14],MAf$M[,3],MAf$M[,7],MAf$M[,11],MAf$M[,15],MAf$M[,4],MAf$M[,8],
        MAf$M[,12],MAf$M[,16], ylim=mlim)
boxplot(MAw$M[,1],MAw$M[,5],MAw$M[,9],MAw$M[,13],MAw$M[,2],MAw$M[,6],MAw$M[,10],
        MAw$M[,14],MAw$M[,3],MAw$M[,7],MAw$M[,11],MAw$M[,15],MAw$M[,4],MAw$M[,8],
        MAw$M[,12],MAw$M[,16], ylim=mlim)
```

```
#=============================================================================
# BOXPLOTS OF A VALUES
#-----------------------------------------------------------------------------

boxplot(MAnorm$A[,1],MAnorm$A[,5],MAnorm$A[,9],MAnorm$A[,13],MAnorm$A[,2],
        MAnorm$A[,6],MAnorm$A[,10],MAnorm$A[,14],MAnorm$A[,3],MAnorm$A[,7],
        MAnorm$A[,11],MAnorm$A[,15],MAnorm$A[,4],MAnorm$A[,8],MAnorm$A[,12],
        MAnorm$A[,16], ylim=alim)
boxplot(MAf$A[,1],MAf$A[,5],MAf$A[,9],MAf$A[,13],MAf$A[,2],MAf$A[,6],MAf$A[,10],
        MAf$A[,14],MAf$A[,3],MAf$A[,7],MAf$A[,11],MAf$A[,15],MAf$A[,4],MAf$A[,8],
        MAf$A[,12],MAf$A[,16], ylim=alim)
boxplot(MAw$A[,1],MAw$A[,5],MAw$A[,9],MAw$A[,13],MAw$A[,2],MAw$A[,6],MAw$A[,10],
        MAw$A[,14],MAw$A[,3],MAw$A[,7],MAw$A[,11],MAw$A[,15],MAw$A[,4],MAw$A[,8],
        MAw$A[,12],MAw$A[,16], ylim=alim)

#=============================================================================
# REMOVE ROWS WITH ANY MISSING VALUES (NA)
#-----------------------------------------------------------------------------

MAnorm.f <- remove.NA(MAw)
RGnorm.f <- RG.MA(MAnorm.f)
RGf      <- RG.MA(MAf)

#=============================================================================
# DETERMINE NUMBER OF NEGATIVE CONTROLS LET POST FILTERING
#-----------------------------------------------------------------------------

Null <- subset(MAnorm.f$M, MAnorm.f$genes$Status=="Null",
               select = c("16HP (B1)"))
Buff <- subset(MAnorm.f$M, MAnorm.f$genes$Status=="Buffer",
               select = c("16HP (B1)"))
Neg  <- subset(MAnorm.f$M, MAnorm.f$genes$Status=="Negative",
               select = c("16HP (B1)"))

length(Null)
length(Buff)
length(Neg)
```

```
##############################################################################
# STEP 6a:      REPLACE WEIGHTS WITH MISSING VALUES (NA)                     #
##############################################################################
# CREATE NEW WEIGHTS MATRIX
#       - Where weights (0) are replaced by NA
#-----------------------------------------------------------------------------

wNA <- MAnorm.f$weights
wNA[MAnorm.f$weights==0] <- NA

#=============================================================================
# MULTIPLY NEW MA (MAnorm.NA) OBJECTS BY MATRIX wNA
#       - If wNA value = 1 object row value remains unchanged
#       - If wNA value = NA object row value becomes NA
#-----------------------------------------------------------------------------

MAnorm.NA <- MAnorm.f
MAnorm.NA$A<-MAnorm.f$A*wNA
MAnorm.NA$M<-MAnorm.f$M*wNA

##############################################################################
# STEP 6b:      MERGE REPLICATES (AVERAGE REPLICATE ROWS)                    #
##############################################################################
# REORDER DATA BASED ON ID (GROUP REPLICATE PROBES TOGETHER)
#-----------------------------------------------------------------------------

IDunsort   <-(as.character(MAnorm.NA$genes$ID))
k          <-order(IDunsort)
IDsort     <- IDunsort[k]
MAnorm.O   <- MAnorm.NA[k,]

#=============================================================================
# CREATE NEW MATRICES (avM AND avA)
#-----------------------------------------------------------------------------

avM <- matrix(0,ncol=16,
              nrow=length(unlist(lapply(split(MAnorm.O$M[,1],IDsort),mean,
              na.rm=TRUE))))

avA <- matrix(0,ncol=16,
              nrow=length(unlist(lapply(split(MAnorm.O$A[,1],IDsort),mean,
              na.rm=TRUE))))

#=============================================================================
# AVERAGE REPLICATES FOR M VALUES A VALUES, IDS AND NAMES OF ALL ARRAYS
#-----------------------------------------------------------------------------

for (q in 1:16)
{
    avM[,q] <- unlist(lapply(split(MAnorm.O$M[,q],IDsort),mean, na.rm=TRUE))
}

for (q in 1:16)
{
    avA[,q] <- unlist(lapply(split(MAnorm.O$A[,q],IDsort),mean, na.rm=TRUE))
}

avID    <- names(lapply(split(MAnorm.O$M[,1],IDsort),mean, na.rm=TRUE))

Name    <-(as.character(MAnorm.NA$genes$Name))
Gene    <- cbind(IDunsort, Name)
GM      <- matrix(Gene, nrow = length(IDsort), ncol = 2)
avName  <- getGENEID(avID, GM, idCol = 2)
```

```
#=============================================================================
# CREATE MATRIX OF IDS AND NAMES
#-----------------------------------------------------------------------------

avGene  <- cbind(avID, avName)
avGM    <- matrix(avGene, nrow = length(avID), ncol = 2)

#=============================================================================
# REANNOTATE MATRIX WITH COLUMN NAMES
#-----------------------------------------------------------------------------

conditions      <- colnames(MAnorm.O$M, do.NULL = FALSE)
identity        <- colnames(MAnorm.O$genes[,4:5], do.NULL = FALSE)
colnames(avA)   <- conditions
colnames(avM)   <- conditions
colnames(avGM)  <- identity

#=============================================================================
# CREATE NEW CLASSES: MAlist (MAnorm.a)
#-----------------------------------------------------------------------------

MAnorm.a        <- new("MAList")
MAnorm.a$M      = avM
MAnorm.a$A      = avA
MAnorm.a$genes  = avGM

#############################################################################
#STEP 6c:       IMPUTE MISSING VALUES                                      #
#############################################################################
# REMOVE ROWS WITH ANY MISSING VALUES
#-----------------------------------------------------------------------------

NArmd <- remove.NA(MAnorm.f$M)
NAadd <- NArmd
dim(NArmd)

#=============================================================================
# GENERATE SET PERCENTAGE (percentNA) OF MISSING VALUES
#-----------------------------------------------------------------------------

percentNA     <- 0.06
count         <- percentNA*length(NArmd[,1])

c=1

while (c <= count)
{
    NAadd[as.integer(runif(1)*5996+1), as.integer(runif(1)*16+1)] <- NA
    c=c+1
}

NAtest        <- remove.NA(NAadd)
(dim(NArmd)[1]- dim(NAtest)[1])/dim(NAtest)[1]

#=============================================================================
# IMPUTE DATA WITH INCREASING VALUES OF k
#-----------------------------------------------------------------------------

Ek1     <- ec.knnimp(NAadd, k = 1)
Ek2     <- ec.knnimp(NAadd, k = 2)
Ek3     <- ec.knnimp(NAadd, k = 3)
Ek4     <- ec.knnimp(NAadd, k = 4)
Ek5     <- ec.knnimp(NAadd, k = 5)
Ek6     <- ec.knnimp(NAadd, k = 6)
Ek7     <- ec.knnimp(NAadd, k = 7)
Ek8     <- ec.knnimp(NAadd, k = 8)
Ek9     <- ec.knnimp(NAadd, k = 9)
Ek10    <- ec.knnimp(NAadd, k = 10)
```

```
Ek11    <- ec.knnimp(NAadd, k = 11)
Ek12    <- ec.knnimp(NAadd, k = 12)
Ek13    <- ec.knnimp(NAadd, k = 13)
Ek14    <- ec.knnimp(NAadd, k = 14)
Ek15    <- ec.knnimp(NAadd, k = 15)
Ek16    <- ec.knnimp(NAadd, k = 16)
Ek17    <- ec.knnimp(NAadd, k = 17)
Ek18    <- ec.knnimp(NAadd, k = 18)
Ek19    <- ec.knnimp(NAadd, k = 19)
Ek20    <- ec.knnimp(NAadd, k = 20)
Ek21    <- ec.knnimp(NAadd, k = 21)
Ek22    <- ec.knnimp(NAadd, k = 22)
Ek23    <- ec.knnimp(NAadd, k = 23)
Ek24    <- ec.knnimp(NAadd, k = 24)
Ek25    <- ec.knnimp(NAadd, k = 25)
Ek26    <- ec.knnimp(NAadd, k = 26)
Ek27    <- ec.knnimp(NAadd, k = 27)
Ek28    <- ec.knnimp(NAadd, k = 28)
Ek29    <- ec.knnimp(NAadd, k = 29)
Ek30    <- ec.knnimp(NAadd, k = 30)
Ek40    <- ec.knnimp(NAadd, k = 40)
Ek50    <- ec.knnimp(NAadd, k = 50)
Ek75    <- ec.knnimp(NAadd, k = 75)
Ek100   <- ec.knnimp(NAadd, k = 100)
Ek150   <- ec.knnimp(NAadd, k = 150)
Ek200   <- ec.knnimp(NAadd, k = 200)

#=============================================================================
# CALCULATE NORMALISED ROOT MEAN SQUARE (NRMS) FOR EACH IMPUTATION
#-----------------------------------------------------------------------------

rms1    <- (sqrt(mean((NArmd-Ek1)^2)))/(sqrt(mean(NArmd^2)))
rms2    <- (sqrt(mean((NArmd-Ek2)^2)))/(sqrt(mean(NArmd^2)))
rms3    <- (sqrt(mean((NArmd-Ek3)^2)))/(sqrt(mean(NArmd^2)))
rms4    <- (sqrt(mean((NArmd-Ek4)^2)))/(sqrt(mean(NArmd^2)))
rms5    <- (sqrt(mean((NArmd-Ek5)^2)))/(sqrt(mean(NArmd^2)))
rms6    <- (sqrt(mean((NArmd-Ek6)^2)))/(sqrt(mean(NArmd^2)))
rms7    <- (sqrt(mean((NArmd-Ek7)^2)))/(sqrt(mean(NArmd^2)))
rms8    <- (sqrt(mean((NArmd-Ek8)^2)))/(sqrt(mean(NArmd^2)))
rms9    <- (sqrt(mean((NArmd-Ek9)^2)))/(sqrt(mean(NArmd^2)))
rms10   <- (sqrt(mean((NArmd-Ek10)^2)))/(sqrt(mean(NArmd^2)))
rms11   <- (sqrt(mean((NArmd-Ek11)^2)))/(sqrt(mean(NArmd^2)))
rms12   <- (sqrt(mean((NArmd-Ek12)^2)))/(sqrt(mean(NArmd^2)))
rms13   <- (sqrt(mean((NArmd-Ek13)^2)))/(sqrt(mean(NArmd^2)))
rms14   <- (sqrt(mean((NArmd-Ek14)^2)))/(sqrt(mean(NArmd^2)))
rms15   <- (sqrt(mean((NArmd-Ek15)^2)))/(sqrt(mean(NArmd^2)))
rms16   <- (sqrt(mean((NArmd-Ek16)^2)))/(sqrt(mean(NArmd^2)))
rms17   <- (sqrt(mean((NArmd-Ek17)^2)))/(sqrt(mean(NArmd^2)))
rms18   <- (sqrt(mean((NArmd-Ek18)^2)))/(sqrt(mean(NArmd^2)))
rms19   <- (sqrt(mean((NArmd-Ek19)^2)))/(sqrt(mean(NArmd^2)))
rms20   <- (sqrt(mean((NArmd-Ek20)^2)))/(sqrt(mean(NArmd^2)))
rms21   <- (sqrt(mean((NArmd-Ek21)^2)))/(sqrt(mean(NArmd^2)))
rms22   <- (sqrt(mean((NArmd-Ek22)^2)))/(sqrt(mean(NArmd^2)))
rms23   <- (sqrt(mean((NArmd-Ek23)^2)))/(sqrt(mean(NArmd^2)))
rms24   <- (sqrt(mean((NArmd-Ek24)^2)))/(sqrt(mean(NArmd^2)))
rms25   <- (sqrt(mean((NArmd-Ek25)^2)))/(sqrt(mean(NArmd^2)))
rms26   <- (sqrt(mean((NArmd-Ek26)^2)))/(sqrt(mean(NArmd^2)))
rms27   <- (sqrt(mean((NArmd-Ek27)^2)))/(sqrt(mean(NArmd^2)))
rms28   <- (sqrt(mean((NArmd-Ek28)^2)))/(sqrt(mean(NArmd^2)))
rms29   <- (sqrt(mean((NArmd-Ek29)^2)))/(sqrt(mean(NArmd^2)))
rms30   <- (sqrt(mean((NArmd-Ek30)^2)))/(sqrt(mean(NArmd^2)))
rms40   <- (sqrt(mean((NArmd-Ek40)^2)))/(sqrt(mean(NArmd^2)))
rms50   <- (sqrt(mean((NArmd-Ek50)^2)))/(sqrt(mean(NArmd^2)))
rms75   <- (sqrt(mean((NArmd-Ek75)^2)))/(sqrt(mean(NArmd^2)))
rms100  <- (sqrt(mean((NArmd-Ek100)^2)))/(sqrt(mean(NArmd^2)))
rms150  <- (sqrt(mean((NArmd-Ek150)^2)))/(sqrt(mean(NArmd^2)))
rms200  <- (sqrt(mean((NArmd-Ek200)^2)))/(sqrt(mean(NArmd^2)))
```

```
#=============================================================================
# PLOT NRMS AT DIFFERENT VALUES OF k
#-----------------------------------------------------------------------------
NRMS <- c(rms1,rms2,rms3,rms4,rms5,rms6,rms7,rms8,rms9,rms10,rms11,rms12,rms13,
          rms14,rms15,rms16,rms17,rms18,rms19,rms20,rms21,rms22,rms23,rms24,rms25,
          rms26,rms27,rms28,rms29,rms30,rms40,rms50,rms75,rms100,rms150,rms200)

k     <- c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,
          28,29,30,40,50,75,100,150,200)

NRMS <- c(rms1,rms2,rms3,rms4,rms5,rms6,rms7,rms8,rms9,rms10,rms11,rms12,rms13,
          rms14,rms15,rms16,rms17,rms18,rms19,rms20,rms25,rms30,rms40,rms50,rms75,
          rms100)
k     <- c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,25,30,40,50,75,100)

Hy <- c(rms15,rms16,rms17,rms18,rms19,rms20)
Hx <- c(15,16,17,18,19,20)

plot(k,NRMS, pch=19, cex=1.2, bty="l")
lines(k, NRMS,type="l", ljoin = "round", bty="l",col="black", lwd=2)
points(Hx, Hy, col = "red", pch=19, cex=1)

#=============================================================================
# IMPUTE MISSING VALUES (NA)
#       - Using KNN impute
#       - number of neighbours = 18
#       - default rowmax - allows 50% missing data in rows
#       - default colmax - allows 80% missing data in columns
#       - default maxp - largest block of genes imputed (15000)
#       - default rng.seed - seed used for random number generator (362436069)
#-----------------------------------------------------------------------------

M <- ec.knnimp(MAnorm.a$M, k=18)
A <- ec.knnimp(MAnorm.a$A, k=18)

#=============================================================================
# CREATE NEW CLASS
#       - MAlist (MAnorm.i)
# CREATE NEW OBJECTS FROM IMPUTED M and A DATA
#-----------------------------------------------------------------------------

MAnorm.i   <- MAnorm.a
MAnorm.i$M <- M
MAnorm.i$A <- A

#=============================================================================
# COERCE INTO NEW CLASSES: RGlist (RGnorm.a and RGnorm.i)
#-----------------------------------------------------------------------------

RGnorm.a <- RG.MA(MAnorm.a)
RGnorm.i <- RG.MA(MAnorm.i)

##############################################################################
# STEP 6d.      VISULISE DATA DISTRIBUTIONS THROUGH ALL PREPROCESSING STEPS   #
#               - Density distribution plots                                  #
#               - Boxplots of M values                                        #
#               - Boxplots of A values                                        #
##############################################################################

par(mfrow=c(1,1))

plotDensities(MAnorm.NA)
plotDensities(MAnorm.a)
plotDensities(MAnorm.i)
```

```
boxplot(MAnorm.NA$M[,1],MAnorm.NA$M[,5],MAnorm.NA$M[,9],MAnorm.NA$M[,13],
        MAnorm.NA$M[,2],MAnorm.NA$M[,6],MAnorm.NA$M[,10],MAnorm.NA$M[,14],
        MAnorm.NA$M[,3],MAnorm.NA$M[,7],MAnorm.NA$M[,11],MAnorm.NA$M[,15],
        MAnorm.NA$M[,4],MAnorm.NA$M[,8],MAnorm.NA$M[,12],MAnorm.NA$M[,16],
        ylim=mlim)
boxplot(MAnorm.a$M[,1],MAnorm.a$M[,5],MAnorm.a$M[,9],MAnorm.a$M[,13],
        MAnorm.a$M[,2],MAnorm.a$M[,6],MAnorm.a$M[,10],MAnorm.a$M[,14],
        MAnorm.a$M[,3],MAnorm.a$M[,7],MAnorm.a$M[,11],MAnorm.a$M[,15],
        MAnorm.a$M[,4],MAnorm.a$M[,8],MAnorm.a$M[,12],MAnorm.a$M[,16],ylim=mlim)
boxplot(MAnorm.i$M[,1],MAnorm.i$M[,5],MAnorm.i$M[,9],MAnorm.i$M[,13],
        MAnorm.i$M[,2],MAnorm.i$M[,6],MAnorm.i$M[,10],MAnorm.i$M[,14],
        MAnorm.i$M[,3],MAnorm.i$M[,7],MAnorm.i$M[,11],MAnorm.i$M[,15],
        MAnorm.i$M[,4],MAnorm.i$M[,8],MAnorm.i$M[,12],MAnorm.i$M[,16], ylim=mlim)

boxplot(MAnorm.NA$A[,1],MAnorm.NA$A[,5],MAnorm.NA$A[,9],MAnorm.NA$A[,13],
        MAnorm.NA$A[,2],MAnorm.NA$A[,6],MAnorm.NA$A[,10],MAnorm.NA$A[,14],
        MAnorm.NA$A[,3],MAnorm.NA$A[,7],MAnorm.NA$A[,11],MAnorm.NA$A[,15],
        MAnorm.NA$A[,4],MAnorm.NA$A[,8],MAnorm.NA$A[,12],MAnorm.NA$A[,16],
        ylim=alim)
boxplot(MAnorm.a$A[,1],MAnorm.a$A[,5],MAnorm.a$A[,9],MAnorm.a$A[,13],
        MAnorm.a$A[,2],MAnorm.a$A[,6],MAnorm.a$A[,10],MAnorm.a$A[,14],
        MAnorm.a$A[,3],MAnorm.a$A[,7],MAnorm.a$A[,11],MAnorm.a$A[,15],
        MAnorm.a$A[,4],MAnorm.a$A[,8],MAnorm.a$A[,12],MAnorm.a$A[,16],ylim=alim)
boxplot(MAnorm.i$A[,1],MAnorm.i$A[,5],MAnorm.i$A[,9],MAnorm.i$A[,13],
        MAnorm.i$A[,2],MAnorm.i$A[,6],MAnorm.i$A[,10],MAnorm.i$A[,14],
        MAnorm.i$A[,3],MAnorm.i$A[,7],MAnorm.i$A[,11],MAnorm.i$A[,15],
        MAnorm.i$A[,4],MAnorm.i$A[,8],MAnorm.i$A[,12],MAnorm.i$A[,16],ylim=alim)

##############################################################################
# STEP 7a:      BATCH CORRECTION - COMBAT                                    #
##############################################################################
# READ R CODE FROM FILE (COMBAT.r)
#-----------------------------------------------------------------------------

source('ComBat.R')

#=============================================================================
# CREATE ANNOTATED MATRIX OF M VALUES AND A VALUES
#-----------------------------------------------------------------------------

dataM <- cbind(MAnorm.i$genes,MAnorm.i$M)
dataM[,3:18] <- MAnorm.i$M[,1:16]

dataA <- cbind(MAnorm.i$genes,MAnorm.i$A)
dataA[,3:18] <- MAnorm.i$A[,1:16]

#=============================================================================
# WRITE MATRICS TO R DATA DIRECTORY FOR COMBAT TO READ IN
#-----------------------------------------------------------------------------

write.table(dataM, "dataM.txt", row.names = FALSE, sep="\t")
write.table(dataA, "dataA.txt", row.names = FALSE, sep="\t")

#=============================================================================
# RUN COMBAT FUNCTION ON M VALUES AND A VALUES
#-----------------------------------------------------------------------------

bcM <- ComBat("dataM.txt", "Batch.txt", skip = 2, write=F)
bcA <- ComBat("dataA.txt", "Batch.txt", skip = 2, write=F)

#=============================================================================
# CREATE NEW CLASS: MAlist (MAnorm.bc)
#-----------------------------------------------------------------------------

MAnorm.bc <- MAnorm.i
MAnorm.bc$M <- bcM[,3:18]
MAnorm.bc$A <- bcA[,3:18]
RGnorm.bc <- RG.MA(MAnorm.bc)
```

```
#############################################################################
# STEP 7b:      VISULISE DATA                                               #
#############################################################################
# DATA DISTRIBUTIONS
#       - Density distribution plots
#       - Boxplots of M values
#       - Boxplots of A values
#----------------------------------------------------------------------------

par(mfrow=c(1,1))
plotDensities(MAnorm.bc)
boxplot(MAnorm.bc$M[,1],MAnorm.bc$M[,5],MAnorm.bc$M[,9],MAnorm.bc$M[,13],
        MAnorm.bc$M[,2],MAnorm.bc$M[,6],MAnorm.bc$M[,10],MAnorm.bc$M[,14],
        MAnorm.bc$M[,3],MAnorm.bc$M[,7],MAnorm.bc$M[,11],MAnorm.bc$M[,15],
        MAnorm.bc$M[,4],MAnorm.bc$M[,8],MAnorm.bc$M[,12],MAnorm.bc$M[,16],
        ylim=mlim)
boxplot(MAnorm.bc$A[,1],MAnorm.bc$A[,5],MAnorm.bc$A[,9],MAnorm.bc$A[,13],
        MAnorm.bc$A[,2],MAnorm.bc$A[,6],MAnorm.bc$A[,10],MAnorm.bc$A[,14],
        MAnorm.bc$A[,3],MAnorm.bc$A[,7],MAnorm.bc$A[,11],MAnorm.bc$A[,15],
        MAnorm.bc$A[,4],MAnorm.bc$A[,8],MAnorm.bc$A[,12],MAnorm.bc$A[,16],
        ylim=alim)

#============================================================================
# VISULISE STANDARD DEVIATION OF CONDITIONS BEFORE AND AFTER BATCH CORRECTION
#       - Before  (MAnorm.i)
#       - After   (MAnorm.bc)
#----------------------------------------------------------------------------

par(mfrow=c(1,1))

MHP16i  <-sd(t(MAnorm.i$M[,1:4]))
MHP17i  <-sd(t(MAnorm.i$M[,5:8]))
MFP16i  <-sd(t(MAnorm.i$M[,9:12]))
MFP17i  <-sd(t(MAnorm.i$M[,13:16]))

MHP16bc <-sd(t(MAnorm.bc$M[,1:4]))
MHP17bc <-sd(t(MAnorm.bc$M[,5:8]))
MFP16bc <-sd(t(MAnorm.bc$M[,9:12]))
MFP17bc <-sd(t(MAnorm.bc$M[,13:16]))

AHP16i  <-sd(t(MAnorm.i$A[,1:4]))
AHP17i  <-sd(t(MAnorm.i$A[,5:8]))
AFP16i  <-sd(t(MAnorm.i$A[,9:12]))
AFP17i  <-sd(t(MAnorm.i$A[,13:16]))

AHP16bc <-sd(t(MAnorm.bc$A[,1:4]))
AHP17bc <-sd(t(MAnorm.bc$A[,5:8]))
AFP16bc <-sd(t(MAnorm.bc$A[,9:12]))
AFP17bc <-sd(t(MAnorm.bc$A[,13:16]))

MOriginalSD <- c(MHP16i,MHP17i,MFP16i,MFP17i)
MBatchCorrectedSD <- c(MHP16bc,MHP17bc,MFP16bc,MFP17bc)

AOriginalSD <- c(AHP16i,AHP17i,AFP16i,AFP17i)
ABatchCorrectedSD <- c(AHP16bc,AHP17bc,AFP16bc,AFP17bc)

plot(MOriginalSD,MBatchCorrectedSD, pch=19,bty="l", xlim=c(0,2.5),ylim=c(0,2.5))
abline(0,1, col="red", lty="longdash", lwd=3)
abline(lsfit(MOriginalSD,MBatchCorrectedSD, intercept=FALSE,), col = "blue",
       lwd=2.5)

plot(AOriginalSD,ABatchCorrectedSD, pch=19, bty="l",xlim=c(0,3.5),ylim=c(0,3.5))
abline(0,1, col="red", lty="longdash", lwd=3)
abline(lsfit(AOriginalSD,ABatchCorrectedSD, intercept=FALSE,), col = "blue",
       lwd=2.5)
```

```
#=============================================================================
# VISULISE STANDARD DEVIATION BETWEEN CONDITIONS BEFORE AND AFTER BATCH CORRECTION
#       - Before  (MAnorm.i)
#       - After   (MAnorm.bc)
#-----------------------------------------------------------------------------

rA <-sd(t(cbind(rowMeans(MAnorm.i$A[,1:4], dims = 1),
                rowMeans(MAnorm.i$A[,5:8], dims = 1),
                rowMeans(MAnorm.i$A[,9:12], dims = 1),
                rowMeans(MAnorm.i$A[,13:16], dims = 1))))

rM <-sd(t(cbind(rowMeans(MAnorm.i$M[,1:4], dims = 1),
                rowMeans(MAnorm.i$M[,5:8], dims = 1),
                rowMeans(MAnorm.i$M[,9:12], dims = 1),
                rowMeans(MAnorm.i$M[,13:16], dims = 1))))

cA <-sd(t(cbind(rowMeans(MAnorm.bc$A[,1:4], dims = 1),
                rowMeans(MAnorm.bc$A[,5:8], dims = 1),
                rowMeans(MAnorm.bc$A[,9:12], dims = 1),
                rowMeans(MAnorm.bc$A[,13:16], dims = 1))))

cM <-sd(t(cbind(rowMeans(MAnorm.bc$M[,1:4], dims = 1),
                rowMeans(MAnorm.bc$M[,5:8], dims = 1),
                rowMeans(MAnorm.bc$M[,9:12], dims = 1),
                rowMeans(MAnorm.bc$M[,13:16], dims = 1))))

plot(rM,cM,pch=19,bty="l", xlim=c(0,2.5),ylim=c(0,2.5) )
abline(0,1, col="red", lty="longdash", lwd=3)
abline(lsfit(rM,cM, intercept=FALSE,), col = "blue", lwd=2.5)

plot(rA,cA,pch=19,bty="l", xlim=c(0,2.5),ylim=c(0,2.5))
abline(0,1, col="red", lty="longdash", lwd=3)
abline(lsfit(rA,cA, intercept=FALSE,), col = "blue", lwd=2.5)

#############################################################################
# STEP 8:       FILTER FLAT PATTERNS BY STANDARD DEVIATION                   #
#############################################################################
# CALCULATE SD VECTOR FOR ALL ROWS
#-----------------------------------------------------------------------------
par(mfrow=c(1,1))

sdAll       <-sd(t(MAnorm.bc$M))
meanAll     <- rowMeans(MAnorm.bc$M)
absmeanAll  <- abs(rowMeans(MAnorm.bc$M))

Status1     <-(as.character(MAnorm.f$genes$Status))
ID1         <-(as.character(MAnorm.f$genes$ID))
ID2         <-(as.character(MAnorm.bc$genes[,1]))

IS          <- cbind(ID1,Status1)
ISM         <- matrix(IS, nrow = length(ID1), ncol = 2)
Status2     <- getGENEID(ID2, ISM, idCol = 2)

msd         <- cbind(Status2, sdAll, absmeanAll)
msdM        <- matrix(msd, nrow = length(sdAll), ncol = 3)
colnames(msdM) <- c("Status", "SD", "M")
dim(msd)

hist(sdAll)
abline(v=0.2, col="red", lty = "longdash", lwd=2.5)
hist(abs(meanAll))
abline(v=0.6, col="red", lty = "longdash", lwd=2.5)
summary(sdAll)

sdC <- subset(msdM, msdM[,1]=="Control", select = c("SD","M"))
sdF <- subset(msdM, msdM[,2]<=0.189, select = c("SD","M"))
sdmF <- subset(sdF, sdF[,2]<0.6, select = c("SD","M"))
```

```
plot(absmeanAll,sdAll, pch=19, bty="l")
points(sdF[,2], sdF[,1], col= "grey", pch=19)
points(sdC[,2], sdC[,1], col= "yellow", pch=19, cex=1.2)
abline(h=0.189, col="red", lty = "longdash", lwd=3)

vB <- sdAll

#=============================================================================
# SET LIMITS FOR SD AND MEAN CUTOFF
#-----------------------------------------------------------------------------

sdL = 0.189

#=============================================================================
# SET MAXIMUM ROW NUMBER
#-----------------------------------------------------------------------------

max3 <- length(MAnorm.bc$M[,1])

#=============================================================================
# TEST IF VECTOR sdAll IS SMALLER THAN sdL AND CREATE VECTOR sdB
#       - If sdAll value <= sdL make sdB value = NA
#       - If sdAll value > sdL make sdB value = 1
#-----------------------------------------------------------------------------

n=1
while(n<=(max3))
{

    if  (sdAll[n]<=sdL)
    {
        vB[n]=NA
    }
    else (vB[n]=1)
    n=n+1
}

#=============================================================================
# CREATE NEW MAList AND MULTIPLY OBJECTS BY VECTOR sdB
#       - If sdB value = 1 object row value remains unchanged
#       - If sdB value = NA object row value becomes NA
#-----------------------------------------------------------------------------

MAnorm.v    <- MAnorm.bc
MAnorm.v$A  <- MAnorm.bc$A*vB
MAnorm.v$M  <- MAnorm.bc$M*vB

#=============================================================================
# REMOVE ROWS WITH ANY MISSING VALUES (NA)
#-----------------------------------------------------------------------------

MAdata <- remove.NA(MAnorm.v)
RGdata <- RG.MA(MAdata)

length(MAdata$M[,1])

#=============================================================================
# STEP 8b:      VISULISE DATA DISTRIBUTIONS AFTER QUALITY FILTERING
#               - Density distribution plots
#               - Boxplots of M values
#               - Boxplots of A values
#-----------------------------------------------------------------------------

par(mfrow=c(1,1))

plotDensities(MAdata)
```

```
boxplot(MAdata$M[,1],MAdata$M[,5],MAdata$M[,9],MAdata$M[,13],MAdata$M[,2],
        MAdata$M[,6],MAdata$M[,10],MAdata$M[,14],MAdata$M[,3],MAdata$M[,7],
        MAdata$M[,11],MAdata$M[,15],MAdata$M[,4],MAdata$M[,8],MAdata$M[,12],
        MAdata$M[,16],ylim=mlim)
boxplot(MAdata$A[,1],MAdata$A[,5],MAdata$A[,9],MAdata$A[,13],MAdata$A[,2],
        MAdata$A[,6],MAdata$A[,10],MAdata$A[,14],MAdata$A[,3],MAdata$A[,7],
        MAdata$A[,11],MAdata$A[,15],MAdata$A[,4],MAdata$A[,8],MAdata$A[,12],
        MAdata$A[,16], ylim=alim)

##############################################################################
# STEP 9:       VISULISE DATA DISTRIBUTIONS THROUGH ALL PREPROCESSING STEPS   #
#               - Density distribution plots                                  #
#               - Boxplots of M values                                        #
#               - Boxplots of A values                                        #
##############################################################################
# GRAPHIC SETTINGS
#       - 1 = black
#       - 2 = red
#       - 3 = green
#       - 4 = blue
#-----------------------------------------------------------------------------

par(mfrow=c(1,1))
dye        <- c("57","69", "70", "65", "67", "68", "77", "55", "79", "64", "73",
                "61", "67", "68", "77", "55")
tissue     <- c("16HP", "16HP", "16HP", "16HP", "17HP", "17HP", "17HP", "17HP",
                "16FP", "16FP", "16FP", "16FP", "17FP", "17FP", "17FP", "17FP")
scanner    <- c("780-530 3.411.95","752-660 4.031.96","680-675 41.96",
                "770-670 3.651.9","X","X","X","X","830-619 4.081.98",
                "723-637 3.991.95","730-708 4.181.98","720-640 3.941.93",
                "X","X","X","X")
condition  <- c(1,1,1,1,2,2,2,2,3,3,3,3,4,4,4,4)
batch      <- c(1,2,3,4,1,2,3,4,1,2,3,4,1,2,3,4)

#=============================================================================
# PCA BEFORE AND AFTER BATCH CORRECTION
#       - Before  (MAnorm.i)
#       - After   (MAnorm.bc)
#-----------------------------------------------------------------------------

par(mfrow=c(1,1))

Raw  <- remove.NA(MAd$M)
Rpca <- prcomp(t(Raw), scale.=TRUE)
plot(Rpca$x[,1:2], pch=19, cex=1.5, col= batch, xlim=c(-400, 400),
     ylim=c(-200,200))
text(Rpca$x[,1:2], labels=tissue, col=batch, font=2, cex=2, adj=0)

BC   <- remove.NA(MAbc$M)
Cpca <- prcomp(t(BC), scale.=TRUE)
plot(Cpca$x[,1:2], pch=19, cex=1.5, col= batch, xlim=c(-400, 400),
     ylim=c(-200,200))
text(Cpca$x[,1:2], labels=tissue, col=batch, font=2, cex=2, adj=0)

Norm <- remove.NA(MAnorm$M)
Npca <- prcomp(t(Norm), scale.=TRUE)
plot(Npca$x[,1:2], pch=19, cex=1.5, col= batch, xlim=c(-400, 400),
     ylim=c(-200,200))
text(Npca$x[,1:2], labels=tissue, col=batch, font=2, cex=2, adj=0)

Fpca <- prcomp(t(MAnorm.f$M), scale.=TRUE)
plot(Fpca$x[,1:2], pch=19, cex=1.5, col= batch, xlim=c(-400, 400),
     ylim=c(-200,200))
text(Fpca$x[,1:2], labels=tissue, col=batch, font=2, cex=2, adj=0)

Bpca <- prcomp(t(MAnorm.bc$M), scale.=TRUE)
plot(Bpca$x[,1:2], pch=19, cex=1.5, col= batch, xlim=c(-400, 400),
     ylim=c(-200,200))
```

```
text(Bpca$x[,1:2], labels=tissue, col=batch, font=2, cex=2, adj=0)

Dpca <- prcomp(t(MAdata$M), scale.=TRUE)
plot(Dpca$x[,1:2], pch=19, cex=1.5, col= batch, xlim=c(-400, 400),
     ylim=c(-200,200))
text(Dpca$x[,1:2], labels=tissue, col=batch, font=2, cex=2, adj=0)


par(mfrow=c(1,1))

screeplot(Rpca, npc = 16, ylim=c(0, 30000), col="white")
screeplot(Cpca, npc = 16, ylim=c(0, 20000), col="white")
screeplot(Npca, npc = 16, ylim=c(0, 20000), col="white")
screeplot(Fpca, npc = 16, ylim=c(0, 20000), col="white")
screeplot(Bpca, npc = 16, ylim=c(0, 20000), col="white")
screeplot(Dpca, npc = 16 ,ylim=c(0, 20000), col="white")

summary(Rpca)
summary(Cpca)
summary(Npca)
summary(Fpca)
summary(Bpca)
summary(Dpca)


RGdN        <- RG.MA(remove.NA(MAd))
RGbcN       <- RG.MA(remove.NA(MAbc))
RGnormN     <- RG.MA(remove.NA(MAnorm))
RGnorm.fN   <- RG.MA(remove.NA(MAnorm.f))
RGnorm.bcN  <- RG.MA(remove.NA(MAnorm.bc))
RGdataN     <- RG.MA(remove.NA(MAdata))

par(mfrow=c(1,1))

source("yasmaAdapt.R")

rg.cor(RGdN, method="pearson", cluster= "correlation",
       clustering.method="complete")
rg.cor(RGbcN, method="pearson", cluster= "correlation",
       clustering.method="complete")
rg.cor(RGnormN, method="pearson", cluster= "correlation",
       clustering.method="complete")
rg.cor(RGnorm.fN, method="pearson", cluster= "correlation",
       clustering.method="complete")
rg.cor(RGnorm.bcN, method="pearson", cluster= "correlation",
       clustering.method="complete")
rg.cor(RGdataN, method="pearson", cluster= "correlation",
       clustering.method="complete")

rg.rsq(RGd)
rg.rsq(RGbcN)
rg.rsq(RGnormN)
rg.rsq(RGnorm.fN)
rg.rsq(RGnorm.bcN)
rg.rsq(RGdataN)

R2     <- c(rg.rsq(RGdN),rg.rsq(RGbcN),rg.rsq(RGnormN),rg.rsq(RGnorm.fN),
            rg.rsq(RGnorm.bcN),rg.rsq(RGdataN))
steps  <- c(1,2,3,4,5,6)
barplot(R2, space = 0.5, col="white", names.arg = steps, ylab="R2", ylim=c(0,1),
        bty = "l")
```

```
#############################################################################
# STEP 10:      DIFFERENTIAL TESTING                                        #
#############################################################################
# STEP 10a. BAT-MOUSE COMPARISON                                            #
#############################################################################
# CREATE SEPERATE MATRICES FOR EACH BAT-MOUSE COMPARISON
#---------------------------------------------------------------------------

MA16hp <- MAdata[,1:4]
MA17hp <- MAdata[,5:8]
MA16fp <- MAdata[,9:12]
MA17fp <- MAdata[,13:16]

data16hp <-as.matrix(MA16hp$M)
data17hp <-as.matrix(MA17hp$M)
data16fp <-as.matrix(MA16fp$M)
data17fp <-as.matrix(MA17fp$M)

#===========================================================================
# DEFINE COMPARISON (ONE-SAMPLE)
#---------------------------------------------------------------------------

comp1 <- c(0,0,0,0)

#===========================================================================
# DEDS TEST
#---------------------------------------------------------------------------

deds16hp <- deds.stat.linkC(data16hp,comp1,B=0,tests = c("modt", "fc", "sam"),
                            tail = "abs", extras = c(1, 1, 0.3),
                            distance = c("weuclid"), adj = "fdr",
                            nsig = length(MAdata$M[,1]))
deds17hp <- deds.stat.linkC(data17hp,comp1,B=0,tests = c("modt", "fc", "sam"),
                            tail = "abs", extras = c(1, 1, 0.3),
                            distance = c("weuclid"), adj = "fdr",
                            nsig = length(MAdata$M[,1]))
deds16fp <- deds.stat.linkC(data16fp,comp1,B=0,tests = c("modt", "fc", "sam"),
                            tail = "abs", extras = c(1, 1, 0.3),
                            distance = c("weuclid"), adj = "fdr",
                            nsig = length(MAdata$M[,1]))
deds17fp <- deds.stat.linkC(data17fp,comp1,B=0,tests = c("modt", "fc", "sam"),
                            tail = "abs", extras = c(1, 1, 0.3),
                            distance = c("weuclid"), adj = "fdr",
                            nsig = length(MAdata$M[,1]))

#===========================================================================
# DEFINE TOP GENES IN A TABLE
#---------------------------------------------------------------------------

HP16 <- topgenes(deds16hp, number = length(MAdata$M[,1]),
                 genelist = MA17hp$genes[,1])
HP17 <- topgenes(deds17hp, number = length(MAdata$M[,1]),
                 genelist = MA17hp$genes[,1])
FP16 <- topgenes(deds16fp, number = length(MAdata$M[,1]),
                 genelist = MA17hp$genes[,1])
FP17 <- topgenes(deds17fp, number = length(MAdata$M[,1]),
                 genelist = MA17hp$genes[,1])

write.table(HP16, "BBdedsHP16.txt", row.names = FALSE, sep="\t")
write.table(HP17, "BBdedsHP17.txt", row.names = FALSE, sep="\t")
write.table(FP16, "BBdedsFP16.txt", row.names = FALSE, sep="\t")
write.table(FP17, "BBdedsFP17.txt", row.names = FALSE, sep="\t")

HP16[1:100,]
HP17[1:10,]
FP16[1:10,]
FP17[1:10,]
```

```
HP16sig <- topgenes(deds16hp, number = sum(deds16hp$p <=0.05))
HP17sig <- topgenes(deds17hp, number = sum(deds17hp$p <=0.05))
FP16sig <- topgenes(deds16fp, number = sum(deds16fp$p <=0.05))
FP17sig <- topgenes(deds17fp, number = sum(deds17fp$p <=0.05))

HP16up[1:10,]
HP17up[1:10,]
FP16up[1:10,]
FP17up[1:10,]

#=============================================================================
# NUMBER OF GENES SIGNIFICANT IN EACH COMPARISON
# (at the 0.01 and 0.05 level of significance)
#-----------------------------------------------------------------------------

sum(deds16hp$p <=0.01)
sum(deds16hp$p <=0.05)
sum(deds17hp$p <=0.01)
sum(deds17hp$p <=0.05)
sum(deds16fp$p <=0.01)
sum(deds16fp$p <=0.05)
sum(deds17fp$p <=0.01)
sum(deds17fp$p <=0.05)

sum(deds16hp$p <=0.01 & deds16hp$stats[,3] <=-1)
sum(deds16hp$p <=0.01 & deds16hp$stats[,3] >= 1)
sum(deds17hp$p <=0.01 & deds16hp$stats[,3] <=-1)
sum(deds17hp$p <=0.01 & deds16hp$stats[,3] >= 1)
sum(deds16fp$p <=0.01 & deds16hp$stats[,3] <=-1)
sum(deds16fp$p <=0.01 & deds16hp$stats[,3] >= 1)
sum(deds17fp$p <=0.01 & deds16hp$stats[,3] <=-1)
sum(deds17fp$p <=0.01 & deds16hp$stats[,3] >= 1)

#=============================================================================
# PLOT QUANTILE-QUANTILE PLOTS OF EACH TEST
# (show genes that are significant at adjusted pval <= thresh
#-----------------------------------------------------------------------------

pairs(deds16hp, subset = c(1:length(MAdata$M[,1])), thresh = 0.01, legend = F)
pairs(deds17hp, subset = c(1:length(MAdata$M[,1])), thresh = 0.01, legend = F)
pairs(deds16fp, subset = c(1:length(MAdata$M[,1])), thresh = 0.01, legend = F)
pairs(deds17fp, subset = c(1:length(MAdata$M[,1])), thresh = 0.01, legend = F)

qqnorm(deds16hp, subset = c(1:length(MAdata$M[,1])), thresh = 0.01)
qqnorm(deds17hp, subset = c(1:length(MAdata$M[,1])), thresh = 0.01)
qqnorm(deds16fp, subset = c(1:length(MAdata$M[,1])), thresh = 0.01)
qqnorm(deds17fp, subset = c(1:length(MAdata$M[,1])), thresh = 0.01)

#=============================================================================
# CREATE VECTORS OF P and LFC VALUES FOR EACH COMPARISON AND ORDER THEM BY
# geneOrder COLUMN
#-----------------------------------------------------------------------------

us16hp       <- HP16[,2]
a            <- order(us16hp)
uspval16hp   <- HP16[,3]
uslfc16hp    <- HP16[,5]
spval16hp    <- uspval16hp[a]
slfc16hp     <- uslfc16hp[a]

us17hp       <- HP17[,2]
b            <- order(us17hp)
uspval17hp   <- HP17[,3]
uslfc17hp    <- HP17[,5]
spval17hp    <- uspval17hp[b]
slfc17hp     <- uslfc17hp[b]

us16fp       <- FP16[,2]
```

```
c             <- order(us16fp)
uspval16fp    <- FP16[,3]
uslfc16fp     <- FP16[,5]
spval16fp     <- uspval16fp[c]
slfc16fp      <- uslfc16fp[c]

us17fp        <- FP17[,2]
d             <- order(us17fp)
uspval17fp    <- FP17[,3]
uslfc17fp     <- FP17[,5]
spval17fp     <- uspval17fp[d]
slfc17fp      <- uslfc17fp[d]

usN    <- FP17[,1]
sN     <- usN[d]

#=============================================================================
# CREATE MATRIX OF ORDERED P AND LFC VALUES WITH ASSIGNED COLUMN NAMES
#-----------------------------------------------------------------------------

pVals  <- cbind(sN, spval16hp, spval17hp, spval16fp, spval17fp)
lfc    <- cbind(sN, slfc16hp, slfc17hp, slfc16fp, slfc17fp)

colnames(pVals)      <-c("Name","HP16", "HP17", "FP16", "FP17")
colnames(lfc)        <-c("Name","HP16", "HP17", "FP16", "FP17")

#=============================================================================
# DEFINE LEVEL OF SIGNIFICANCE FOR BINARY SIGNIFICANCE MATRIX
#-----------------------------------------------------------------------------

sig    <- 0.05
fold   <-0
foldup <- 0
folddn <-  0
row    <- length(MAdata$M[,1])
col    <- 5

#=============================================================================
# TEST IF VALUES IN MATRIX pVals ARE LARGER OR SMALLER THAN sig
# AND IF GENES ARE UP OR DOWN -REGULATED
# (If pval value < sig make sigB value = 1)
# (If pval value > sig make sigB value = 0)
# (If lfc value < foldup make sigB value = +ve)
# (If lfc value > folddn make sigB value = -ve)
#-----------------------------------------------------------------------------

binary <- pVals

j=2

while(j<=(col))
{
      i=1
      while(i<=(row))
      {

              if (pVals[i,j]>sig)
              {
                    binary[i,j]=0
              }
              else
              {
                    binary[i,j]=1
              }
             i=i+1
      }
      j=j+1
}
```

```
binary[1:5,]

logic <- binary

j=2

while(j<=(col))
{
       i=1

       while(i<=(row))
       {
           if (lfc[i,j]>fold)
           {
              logic[i,j]=as.numeric(logic[i,j])*-1
           }
           else
           {
              logic[i,j]=as.numeric(logic[i,j])*1
           }
           i=i+1
       }
       j=j+1
}

j=2

while(j<=(col))
{
       i=1
       while(i<=(row))
       {
           if (lfc[i,j]>foldup & lfc[i,j]<folddn )
           {
              logic[i,j]=0
           }
           i=i+1
       }
       j=j+1
}

#=============================================================================
# SPLIT BINARY MATRIX logic INTO A MATRIX FOR EACH COMPARISON GROUP
#-----------------------------------------------------------------------------

sig16 <- cbind(logic[,2],logic[,4])
colnames(sig16 ) <- c("HP16", "FP16")

sig17 <- cbind(logic[,3],logic[,5])
colnames(sig17 ) <- c("HP17", "FP17")


#=============================================================================
# CALCULATE VENN COUNTS FOR EACH GROUP
#-----------------------------------------------------------------------------

up16   <- vennCounts(sig16, include="up")
down16 <- vennCounts(sig16, include="down")
up17   <- vennCounts(sig17, include="up")
down17 <- vennCounts(sig17, include="down")

up16
down16
up17
down17
```

```
#=========================================================================
# DRAW VENN DIAGRAMS FOR EACH COMPARISON
#-------------------------------------------------------------------------

nm16 <-c("16HP","16FP")
nm17 <-c("17HP","17FP")

par(mfrow=c(2,1))
vennDiagram(sig16, include=c("up","down"), names = nm16, mar=rep(1,4),
            cex=1.5, lwd=1, counts.col=c("green","red"))
vennDiagram(sig17, include=c("up","down"), names = nm17, mar=rep(1,4),
            cex=1.5, lwd=1, counts.col=c("green","red"))

#=========================================================================
# UPREGULATED IN HP AS OPPOSED TO MOUSE
#-------------------------------------------------------------------------
row = length(MAdata$M[,1])
i=1
upHP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] == 1)&&
              (logic[i,3] == 1)&&
              (logic[i,4] == 0)&&
              (logic[i,5] == 0)  )
              {upHP[i] =as.character(logic[i,1])}
       else  {upHP[i] = NA}
       i=i+1
}

#=========================================================================
# DOWNREGULATED IN HP AS OPPOSED TO MOUSE
#-------------------------------------------------------------------------
i=1
dnHP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] == -1)&&
              (logic[i,3] == -1)&&
              (logic[i,4] ==  0)&&
              (logic[i,5] ==  0)  )
              {dnHP[i] =as.character(logic[i,1])}
       else  {dnHP[i] = NA}
       i=i+1
}

#=========================================================================
# UPREGULATED IN FP AS OPPOSED TO MOUSE
#-------------------------------------------------------------------------
i=1
upFP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] == 0)&&
              (logic[i,3] == 0)&&
              (logic[i,4] == 1)&&
              (logic[i,5] == 1)  )
              {upFP[i] =as.character(logic[i,1])}
       else  {upFP[i] = NA}
       i=i+1
}
```

```
#=============================================================================
# DOWNREGULATED IN FP AS OPPOSED TO MOUSE
#-----------------------------------------------------------------------------
i=1
dnFP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] ==  0)&&
              (logic[i,3] ==  0)&&
              (logic[i,4] == -1)&&
              (logic[i,5] == -1)  )
              {dnFP[i] =as.character(logic[i,1])}
       else  {dnFP[i] = NA}
       i=i+1
}

#=============================================================================
# UPREGULATED IN 16 AS OPPOSED TO MOUSE
#-----------------------------------------------------------------------------

i=1
up16 <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] == 1)&&
              (logic[i,3] == 0)&&
              (logic[i,4] == 1)&&
              (logic[i,5] == 0)  )
              {up16[i] =as.character(logic[i,1])}
       else  {up16[i] = NA}
       i=i+1
}

#=============================================================================
# DOWNREGULATED IN 16 AS OPPOSED TO MOUSE
#-----------------------------------------------------------------------------
i=1
dn16 <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] == -1)&&
              (logic[i,3] ==  0)&&
              (logic[i,4] == -1)&&
              (logic[i,5] ==  0)  )
              {dn16[i] =as.character(logic[i,1])}
       else  {dn16[i] = NA}
       i=i+1
}

#=============================================================================
# UPREGULATED IN 17 AS OPPOSED TO MOUSE
#-----------------------------------------------------------------------------
i=1
up17 <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] == 0)&&
              (logic[i,3] == 1)&&
              (logic[i,4] == 0)&&
              (logic[i,5] == 1)  )
              {up17[i] =as.character(logic[i,1])}
       else  {up17[i] = NA}
       i=i+1
}
```

```
#=============================================================================
# DOWNREGULATED IN 17 AS OPPOSED TO MOUSE
#-----------------------------------------------------------------------------
i=1
dn17 <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] ==  0)&&
              (logic[i,3] == -1)&&
              (logic[i,4] ==  0)&&
              (logic[i,5] == -1)  )
              {dn17[i] =as.character(logic[i,1])}
       else  {dn17[i] = NA}
       i=i+1
}

#=============================================================================
# UPREGULATED IN 16HP AS OPPOSED TO MOUSE
#-----------------------------------------------------------------------------

i=1
up16HP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] == 1)&&
              (logic[i,3] == 0)&&
              (logic[i,4] == 0)&&
              (logic[i,5] == 0)  )
              {up16HP[i] =as.character(logic[i,1])}
       else  {up16HP[i] = NA}
       i=i+1
}

#=============================================================================
# DOWNREGULATED IN 16HP AS OPPOSED TO MOUSE
#-----------------------------------------------------------------------------
i=1
dn16HP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] == -1)&&
              (logic[i,3] ==  0)&&
              (logic[i,4] ==  0)&&
              (logic[i,5] ==  0)  )
              {dn16HP[i] =as.character(logic[i,1])}
       else  {dn16HP[i] = NA}
       i=i+1
}

#=============================================================================
# UPREGULATED IN 16FP AS OPPOSED TO MOUSE
#-----------------------------------------------------------------------------
i=1
up16FP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] == 0)&&
              (logic[i,3] == 0)&&
              (logic[i,4] == 1)&&
              (logic[i,5] == 0)  )
              {up16FP[i] =as.character(logic[i,1])}
       else  {up16FP[i] = NA}
       i=i+1
}
```

```
#=============================================================================
# DOWNREGULATED IN 16FP AS OPPOSED TO MOUSE
#-----------------------------------------------------------------------------
i=1
dn16FP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] ==  0)&&
              (logic[i,3] ==  0)&&
              (logic[i,4] == -1)&&
              (logic[i,5] ==  0)  )
              {dn16FP[i] =as.character(logic[i,1])}
       else  {dn16FP[i] = NA}
       i=i+1
}

#=============================================================================
# UPREGULATED IN 17HP AS OPPOSED TO MOUSE
#-----------------------------------------------------------------------------
i=1
up17HP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] == 0)&&
              (logic[i,3] == 1)&&
              (logic[i,4] == 0)&&
              (logic[i,5] == 0)  )
              {up17HP[i] =as.character(logic[i,1])}
       else  {up17HP[i] = NA}
       i=i+1
}

#=============================================================================
# DOWNREGULATED IN 17HP AS OPPOSED TO MOUSE
#-----------------------------------------------------------------------------
i=1
dn17HP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] ==  0)&&
              (logic[i,3] == -1)&&
              (logic[i,4] ==  0)&&
              (logic[i,5] ==  0)  )
              {dn17HP[i] =as.character(logic[i,1])}
       else  {dn17HP[i] = NA}
       i=i+1
}

#=============================================================================
# UPREGULATED IN 17FP AS OPPOSED TO MOUSE
#-----------------------------------------------------------------------------
i=1
up17FP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] == 0)&&
              (logic[i,3] == 0)&&
              (logic[i,4] == 0)&&
              (logic[i,5] == 1)  )
              {up17FP[i] =as.character(logic[i,1])}
       else  {up17FP[i] = NA}
       i=i+1
}
```

```
#=============================================================================
# DOWNREGULATED IN 17FP AS OPPOSED TO MOUSE
#-----------------------------------------------------------------------------
i=1
dn17FP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (logic[i,2] ==  0)&&
              (logic[i,3] ==  0)&&
              (logic[i,4] ==  0)&&
              (logic[i,5] == -1)  )
              {dn17FP[i] =as.character(logic[i,1])}
       else  {dn17FP[i] = NA}
       i=i+1
}

#=============================================================================
# REMOVE NA
#-----------------------------------------------------------------------------

upHP <-upHP[!is.na(upHP)]
dnHP <-dnHP[!is.na(dnHP)]
upFP <-upFP[!is.na(upFP)]
dnFP <-dnFP[!is.na(dnFP)]

up16 <-up16[!is.na(up16)]
dn16 <-dn16[!is.na(dn16)]
up17 <-up17[!is.na(up17)]
dn17 <-dn17[!is.na(dn17)]

up16HP <-up16HP[!is.na(up16HP)]
dn16HP <-dn16HP[!is.na(dn16HP)]
up16FP <-up16FP[!is.na(up16FP)]
dn16FP <-dn16FP[!is.na(dn16FP)]

up17HP <-up17HP[!is.na(up17HP)]
dn17HP <-dn17HP[!is.na(dn17HP)]
up17FP <-up17FP[!is.na(up17FP)]
dn17FP <-dn17FP[!is.na(dn17FP)]

#=============================================================================
# NUMBER OF GENES IN EACH GROUP
#-----------------------------------------------------------------------------

length(upHP)
length(dnHP)
length(upFP)
length(dnFP)

length(up16)
length(dn16)
length(up17)
length(dn17)

length(up16HP)
length(dn16HP)
length(up16FP)
length(dn16FP)

length(up17HP)
length(dn17HP)
length(up17FP)
length(dn17FP)
```

```
#=============================================================================
# LIST OF GENES IN EACH GROUP
#-----------------------------------------------------------------------------

upHP
dnHP
upFP
dnFP

up16
dn16
up17
dn17

up16HP
dn16HP
up16FP
dn16FP

up17HP
dn17HP
up17FP
dn17FP

#=============================================================================
# WRITE OUT TABLES OF GENES IN EACH GROUP
#-----------------------------------------------------------------------------

g1  <- data.frame(cbind(Group="upHP", Gene = upHP))
g2  <- data.frame(cbind(Group="dnHP", Gene = dnHP))
g3  <- data.frame(cbind(Group="upFP", Gene = upFP))
g4  <- data.frame(cbind(Group="dnFP", Gene = dnFP))
g5  <- data.frame(cbind(Group="up16", Gene = up16))
g6  <- data.frame(cbind(Group="dn16", Gene = dn16))
g7  <- data.frame(cbind(Group="up17", Gene = up17))
g8  <- data.frame(cbind(Group="dn17", Gene = dn17))
g9  <- data.frame(cbind(Group="up16HP", Gene = up16HP))
g10 <- data.frame(cbind(Group="dn16HP", Gene = dn16HP))
g11 <- data.frame(cbind(Group="up16FP", Gene = up16FP))
g12 <- data.frame(cbind(Group="dn16FP", Gene = dn16FP))
g13 <- data.frame(cbind(Group="up17HP", Gene = up17HP))
g14 <- data.frame(cbind(Group="dn17HP", Gene = dn17HP))
g15 <- data.frame(cbind(Group="up17FP", Gene = up17FP))
g16 <- data.frame(cbind(Group="dn17FP", Gene = dn17FP))

groupsBM <- rbind(g1,g2, g3,g4, g5,g6,g7,g8,g9,g10,g11,g12,g13,g14, g15,g16)

write.table(groupsBM, "BMgenelistTEST.txt", row.names = FALSE, sep="\t")

##############################################################################
# STEP 10b. BAT-BAT COMPARISON                                               #
##############################################################################
# CREATE NEW CLASSES FOR  EACH BAT-BAT (CS16 AND CS17) COMPARISON
#-----------------------------------------------------------------------------

MAdata[1:5,]
MAdata.16           <- MAdata[,1:8]
MAdata.16$M[,5:8]   <- MAdata$M[,9:12]
MAdata.16$A[,5:8]   <- MAdata$A[,9:12]
names16 <- c("16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)", "16FP (B1)",
             "16FP (B2)", "16FP (B3)", "16FP (B4)")
colnames(MAdata.16$M) <- names16
colnames(MAdata.16$A) <- names16
MAdata.16[1:5,]

MAdata.17           <- MAdata[,1:8]
MAdata.17$M[,1:4]   <- MAdata.17$M[,5:8]
MAdata.17$M[,5:8]   <- MAdata$M[,13:16]
```

```
MAdata.17$A[,1:4]   <- MAdata.17$A[,5:8]
MAdata.17$A[,5:8]   <- MAdata$A[,13:16]
names17 <- c("17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)", "17FP (B1)",
             "17FP (B2)", "17FP (B3)", "17FP (B4)")
colnames(MAdata.17$M) <- names17
colnames(MAdata.17$A) <- names17
MAdata.17[1:5,]

#=============================================================================
# CREATE SEPERATE MATRICES FOR EACH BAT-BAT COMPARISON
#-----------------------------------------------------------------------------

data16 <- as.matrix(MAdata.16$M)
data17 <- as.matrix(MAdata.17$M)

data16[1:5,]
data17[1:5,]

#=============================================================================
# DEFINE COMPARISON (TWO-SAMPLE)
#-----------------------------------------------------------------------------

comp2 <- c(0,0,0,0,1,1,1,1)

#=============================================================================
# DEDS TEST
#-----------------------------------------------------------------------------

deds16  <- deds.stat.linkC(data16,comp2,B=0, tests = c("modt", "fc", "sam"),
                           tail = "abs", extras = c(2, 2, 0.3),
                           distance = c("weuclid"), adj="fdr",
                           nsig = length(MAdata.16[,1]))
deds17  <- deds.stat.linkC(data17,comp2,B=0, tests = c("modt", "fc", "sam"),
                           tail = "abs", extras = c(2, 2, 0.3),
                           distance = c("weuclid"), adj="fdr",
                           nsig = length(MAdata.16[,1]))

#=============================================================================
# DEFINE TOP GENES IN A TABLE
#-----------------------------------------------------------------------------

diff16  <- topgenes(deds16, number = length(MAdata.16[,1]),
                    genelist = MAdata.16$genes[,1])
diff17  <- topgenes(deds17, number = length(MAdata.17[,1]),
                    genelist = MAdata.17$genes[,1])

write.table(diff16, "BBdeds16.txt", row.names = FALSE, sep="\t")
write.table(diff17, "BBdeds17.txt", row.names = FALSE, sep="\t")

diff16[1:6,]
diff17[1:15,]

#=============================================================================
# NUMBER OF GENES SIGNIFICANT IN EACH COMPARISON
# (at the 0.01 and 0.05 level of significance)
#-----------------------------------------------------------------------------

sum(deds16$p  <=0.01)
sum(deds16$p  <=0.1)

sum(deds17$p  <=0.01)
sum(deds17$p  <=0.05)
```

```
#=============================================================================
# CREATE TOPTABLES WITH ONLY SIGNIFICANT GENES PRESENT
#-----------------------------------------------------------------------------

diff16  <- topgenes(deds16, number = sum(deds16$p<=0.05),
                    genelist = MAdata.16$genes[,1])
diff17  <- topgenes(deds17, number = sum(deds17$p<=0.05),
                    genelist = MAdata.17$genes[,1])

#=============================================================================
# PLOT QUANTILE-QUANTILE PLOTS OF EACH TEST
# (show genes that are significant at adjusted pval <= thresh
#-----------------------------------------------------------------------------

pairs(deds16, subset = c(1:length(MAdata.16[,1])), thresh = 0.01, legend = F)
pairs(deds17, subset = c(1:length(MAdata.17[,1])), thresh = 0.01, legend = F)

qqnorm(deds16, subset = c(1:(length(MAdata.16[,1])), thresh = 0.01)
qqnorm(deds17, subset = c(1:length(MAdata.17[,1])), thresh = 0.01)

#=============================================================================
# CREATE TOPTABLES WITH TOP 5% GENES PRESENT
#-----------------------------------------------------------------------------

sig16 <- topgenes(deds16, number = (length(deds16$p)*0.05),
                  genelist = MAdata.16$genes[,1])
sig17 <- topgenes(deds17, number = (length(deds16$p)*0.05),
                  genelist = MAdata.17$genes[,1])

#=============================================================================
# COMPARE TOPTABLE GENES BETWEEN STAGES
#-----------------------------------------------------------------------------

gene16        <- as.matrix(sig16[,1])
gene17        <- as.matrix(sig17[,1])
sigAll        <- rbind(sig16,sig17)
geneAll       <- sigAll[,1]
geneAll       <- as.character(geneAll)

g             <-order(geneAll)
geneSort      <- geneAll[g]
order         <- sigAll[g,]

avGene        <- names(lapply(split(order,geneSort),mean, na.rm=TRUE))
length(avGene)

count         <- matrix(0, ncol=2, nrow = length(avGene))
colnames(count) <- c("Sig16", "Sig17")

#=============================================================================
# SET PARAMETERS
#-----------------------------------------------------------------------------

FCup <- -0.263
FCdn <-  0.263

maxA <- length(avGene)
maxB <- length(gene16[,1])
maxC <- length(gene17[,1])
```

```
#=============================================================================
# TEST IF GENES ARE REDUNDENT AMONG LISTS AT A SPECIFIC FOLD CHANGE LEVEL
# (If fold change < FCup make sigB value = +ve)
# (If fold change > FCdn make sigB value = -ve)
#-----------------------------------------------------------------------------
i=1

while(i<=maxA)
{
       j=1
       while(j<=maxB)
       {
           if(avGene[i]==sig16[j,1])
           {
              count[i,1]=sig16[j,5]
           }
           j=j+1
       }
       i=i+1
}


i=1

while(i<=maxA)
{
       k=1
       while(k<=maxC)
       {
           if(avGene[i]==sig17[k,1])
           {
              count[i,2]= sig17[k,5]
           }
           k=k+1
       }
       i=i+1
}

count[1:20,]

logicB <- count
logicB [1:20,]

m=1
while(m<=2)
{
       l=1
       while(l<=(maxA))
       {

           if (logicB[l,m]>FCup && logicB[l,m]< FCdn)
           {
              logicB[l,m]=0
           }
           if (logicB[l,m]<FCup)
           {
              logicB[l,m]=-1
           }

           if (logicB[l,m]>FCdn)
           {
              logicB[l,m]=1
           }

           l=l+1
       }

       m=m+1}
```

```
inv16 <- as.numeric(logicB[,1])*-1
inv17 <- as.numeric(logicB[,2])*-1

logicB[,1] <- inv16
logicB[,2] <- inv17

logicB[1:10,]

matrix              <- matrix(0, ncol=3, nrow = length(avGene))
colnames(matrix)    <- c("Names", "Sig16", "Sig17")
matrix[,2:3]        <- logicB
matrix[,1]          <- avGene
matrix[1:10,]

#=============================================================================
# CALCULATE VENN COUNTS FOR EACH GROUP
#-----------------------------------------------------------------------------

vennCounts(matrix[,2:3], include="up")
vennCounts(matrix[,2:3], include="down")

#=============================================================================
# DRAW VENN DIAGRAM
#-----------------------------------------------------------------------------

par(mfrow=c(1,1))
nm <-c("16","17")
vennDiagram(matrix[,2:3], include=c("up","down"), names = nm, mar=rep(1,4),
            cex=1.5, lwd=1, counts.col=c("green","red"))

#=============================================================================
# UPREGULATED IN 16 and 17 BAT FORELIMBS
#-----------------------------------------------------------------------------

row = maxA
i=1
upBHP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
      if (   (matrix[i,2] == 1)&&
             (matrix[i,3] == 1))

      upBHP[i] =as.character(matrix[i,1])

      else  {upBHP[i] = NA}
      i=i+1
}

#=============================================================================
# DOWNREGULATED IN 16 and 17 BAT FORELIMBS
#-----------------------------------------------------------------------------
row = maxA
i=1
dnBHP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
      if (   (matrix[i,2] == -1)&&
             (matrix[i,3] == -1))

      dnBHP[i] =as.character(matrix[i,1])

      else  {dnBHP[i] = NA}
      i=i+1
}
```

```
#=============================================================================
# UPREGULATED IN 16 BAT FORELIMBS ONLY
#-----------------------------------------------------------------------------
row = maxA
i=1
up16BHP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (matrix[i,2] == 1)&&
              (matrix[i,3] == 0))

       up16BHP[i] =as.character(matrix[i,1])

       else  {up16BHP[i] = NA}
       i=i+1
}

#=============================================================================
# DOWNREGULATED IN 16 BAT FORELIMBS ONLY
#-----------------------------------------------------------------------------
row = maxA
i=1
dn16BHP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (matrix[i,2] == -1)&&
              (matrix[i,3] ==  0))

       dn16BHP[i] =as.character(matrix[i,1])

       else  {dn16BHP[i] = NA}
       i=i+1
}


#=============================================================================
# UPREGULATED IN 17 BAT FORELIMBS ONLY
#-----------------------------------------------------------------------------
row = maxA
i=1
up17BHP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (matrix[i,2] == 0)&&
              (matrix[i,3] == 1))

       up17BHP[i] =as.character(matrix[i,1])

       else  {up17BHP[i] = NA}
       i=i+1
}

#=============================================================================
# DOWNREGULATED IN 17 BAT FORELIMBS ONLY
#-----------------------------------------------------------------------------
row = maxA
i=1
dn17BHP <- vector(mode = "numeric", length = row)
while(i<=(row))
{
       if (   (matrix[i,2] ==  0)&&
              (matrix[i,3] == -1))

       dn17BHP[i] =as.character(matrix[i,1])

       else  {dn17BHP[i] = NA}
       i=i+1
}
```

```
#=============================================================================
# REMOVE NA
#-----------------------------------------------------------------------------

upBHP <-upBHP[!is.na(upBHP)]
dnBHP <-dnBHP[!is.na(dnBHP)]
up16BHP <-up16BHP[!is.na(up16BHP)]
dn16BHP <-dn16BHP[!is.na(dn16BHP)]
up17BHP <-up17BHP[!is.na(up17BHP)]
dn17BHP <-dn17BHP[!is.na(dn17BHP)]

#=============================================================================
# NUMBER OF GENES IN EACH GROUP
#-----------------------------------------------------------------------------

length(upBHP)
length(dnBHP)
length(up16BHP)
length(dn16BHP)
length(up17BHP)
length(dn17BHP)

#=============================================================================
# LIST OF GENES IN EACH GROUP
#-----------------------------------------------------------------------------

upBHP
dnBHP
up16BHP
dn16BHP
up17BHP
dn17BHP

#=============================================================================
# WRITE OUT TABLES OF GENES IN EACH GROUP
#-----------------------------------------------------------------------------

h1  <- data.frame(cbind(Group="upBHP", Gene = upBHP))
h2  <- data.frame(cbind(Group="dnBHP", Gene = dnBHP))
h3  <- data.frame(cbind(Group="up16BHP", Gene = up16BHP))
h4  <- data.frame(cbind(Group="dn16BHP", Gene = dn16BHP))
h5  <- data.frame(cbind(Group="up17BHP", Gene = up17BHP))
h6  <- data.frame(cbind(Group="dn17BHP", Gene = dn17BHP))

groupsBB <- rbind(h1,h2,h3,h4,h5,h6)

write.table(groupsBB, "BBgenelist.txt", row.names = FALSE, sep="\t")

#############################################################################
# STEP 10c. DATA FOR GRAPHING EXPRESSION                                    #
#############################################################################

MAnorm.bc[1:5,]

#=============================================================================
# COERCE INTO A DIFFERENT CLASS: MAlist INTO A RGlist
#-----------------------------------------------------------------------------
RGnorm       <- RG.MA(MAnorm)
RGnorm[1:5,]

mouseExp     <- rowMeans(RGnorm$R, dims = 1)
mouseSD      <- sd(t(RGnorm$R))

HP16Exp      <      - rowMeans(RGnorm$G[,1:4], dims = 1)
HP16SD       <- sd(t(RGnorm$G[,1:4]))

HP17Exp      <- rowMeans(RGnorm$G[,5:8], dims = 1)
HP17SD       <- sd(t(RGnorm$G[,5:8]))
```

```
FP16Exp       <- rowMeans(RGnorm$G[,9:12], dims = 1)
FP16SD        <- sd(t(RGnorm$G[,9:12]))

FP17Exp       <- rowMeans(RGnorm$G[,13:16], dims = 1)
FP17SD        <- sd(t(RGnorm$G[,13:16]))

allExpNF      <- cbind(RGnorm$genes, mouseExp, HP17Exp,  FP17Exp,HP16Exp, FP16Exp,
                       mouseSD,  HP17SD, FP17SD, HP16SD,FP16SD)


allExp[1:5,]

write.table(allExpNF, "allExp no filtering.txt", row.names = FALSE, sep="\t")

#============================================================================
# EXTRACT NAMES OF REPLICATE GENES
#----------------------------------------------------------------------------

RGnorm.b      <- RGnorm.bc
Nunsort       <-(as.character(RGnorm.bc$genes[,2]))
l             <-order(Nunsort)
Nsort         <- Nunsort[l]
RGnorm.b      <- RGnorm.bc[l,]

RGnorm.b[1:5,]

x             <- RGnorm.b$genes[,2]
x[1:5]
dup           <- duplicated(x)
bin           <- dup


summary(dup)
dup[1:100]

max3 = length(dup)
i=1
while(i<=(max3))
{

       if (dup[i]>0)
       {
              bin[i]=1
       }
       else
       {
              bin[i]=NA
       }

       i=i+1
}

summary(bin)
bin[1:5]
RGnorm.rp     <- RGnorm.b
RGnorm.rp$R   <- RGnorm.b$R*bin
RGnorm.rp$G   <- RGnorm.b$G*bin
RGnorm.rp[1:5,]
length(RGnorm.rp$R[,1])
RGnorm.rp     <- remove.NA(RGnorm.rp)
length(RGnorm.rp$R[,1])
RGnorm.rp$genes[1:10,]
RGnorm.x      <- RGnorm.rp
Xunsort       <-(as.character(RGnorm.rp$genes[,2]))
m <-order(Xunsort)
Xsort         <- Xunsort[m]
RGnorm.x      <- RGnorm.rp[m,]
avN <- names(lapply(split(RGnorm.x$R[,1],Xsort),mean, na.rm=TRUE))
```

```
length(avN)
avN
write.table(avN, "replicates", row.names = FALSE, sep="\t")

write.table(MAnorm.bc$genes, "names check batch.txt", row.names = FALSE, sep="\t")

subset(RGnorm$genes,RGnorm$genes$ID=="M400000053", select = c("ID", "Name"))
subset(RGnorm$G, RGnorm$genes$ID=="M400000053",  select =
       c( "16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
          "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
          "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
          "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))

length(RGnorm.bc$R[,1])

subset(RGnorm.bc$genes,RGnorm.bc$genes[,1]=="M200015469",select = c("ID", "Name"))
subset(RGnorm.bc$G, RGnorm.bc$genes[,1]=="M200015469",  select =
       c( "16HP (B1)", "16HP (B2)", "16HP (B3)", "16HP (B4)",
          "17HP (B1)", "17HP (B2)", "17HP (B3)", "17HP (B4)",
          "16FP (B1)", "16FP (B2)", "16FP (B3)", "16FP (B4)",
          "17FP (B1)", "17FP (B2)", "17FP (B3)", "17FP (B4)"))

HP16 <- MAnorm.bc$M[,1:4]
FP16 <- MAnorm.bc$M[,9:12]

HP16 <- rowMeans(HP16)
FP16 <- rowMeans(FP16)

FC <-HP16-FP16
length(MAnorm$M[,1])
CHECK <-cbind(MAnorm.bc$genes, RGnorm.bc$R, RGnorm.bc$G)
write.table(MAdata$genes, "NAMES data.txt", row.names = FALSE, sep="\t")
write.table(MAnorm.bc$genes, "NAMES bc.txt", row.names = FALSE, sep="\t")
write.table(MAnorm$genes, "NAMES all the rest.txt", row.names = FALSE, sep="\t")

write.table(MAnorm.bc$M, "Mvalues.txt", row.names = FALSE, sep="\t")

write.table(Status2, "Gene status.txt", row.names = FALSE, sep="\t")


control<- c(Status2,MAnorm.bc$M)
control[1:5]

length(Status2)

dim(MAnorm.bc$genes)
dim(MAnorm.bc$M)

###############################################################################
```